



UNIVERSIDAD CATÓLICA ANDRÉS BELLO
FACULTAD DE HUMANIDADES Y EDUCACIÓN
ESCUELA DE COMUNICACIÓN SOCIAL
MENCIÓN PERIODISMO
TRABAJO DE GRADO

PROYECTO CIC-BIG DATA

Tesistas:

GIAMBALVO KAUFATI, Estefanía E.

JIMÉNEZ CRUZ, Nancy M.

Tutores:

DELGADO FLORES, Carlos

PÉREZ QUINTERO, José L.

Caracas, septiembre 2015

A mis padres, Giuseppe y Mayra
A la memoria de mi abuela Georgette

Estefanía

AGRADECIMIENTOS Y RECONOCIMIENTOS

El presente trabajo de grado no hubiera sido posible sin la valiosa colaboración del profesor Carlos Delgado Flores, nuestro tutor institucional, quien con mucha paciencia y dedicación supo orientarnos para realizar el proyecto.

Agradecemos al profesor José Luis Pérez Quintero, nuestro tutor académico, por brindarnos su valiosa asesoría cuando más lo necesitábamos.

A nuestros padres, familiares y amigos; gracias por su apoyo incondicional en todo momento.

ÍNDICE

	Páginas
I. INTRODUCCIÓN	7
II. PLANTEAMIENTO DEL PROBLEMA	11
III. MARCO CONTEXTUAL	13
1. Centro de Investigación de la Comunicación de la Universidad Católica Andrés Bello (CIC-UCAB)	13
IV. MARCO TEÓRICO	15
1. Big Data	15
1.1. Definición	15
1.2. Antecedentes	17
1.3. Características	20
1.4. Tipos de datos	22
1.5. Big Data en América Latina	24
2. Open Data	27
2.1. Definición	27
2.2. Características	28
3. Periodismo de datos	30
3.1. Definición	30
3.2. Periodismo de datos como herramienta para la investigación	31
3.3. Periodismo de datos y su uso en Venezuela	33
3.4. Big Data y su aporte al Periodismo de datos	36
4. Walter Benjamin	37
5. Mindomo	39

V. MARCO METODOLÓGICO	40
1. Establecimiento de objetivos	40
1.1. Objetivo general	40
1.2. Objetivos específicos	40
2. Modalidad de tesis	40
3. Tipo y diseño de investigación	41
4. Plan de trabajo de la pasantía	42
VI. ANÁLISIS Y DISCUSIÓN DE LOS RESULTADOS	44
1. Categorización	45
1.1. Construcción de las categorías	46
1.1.1. Géneros	46
1.1.2. Temáticas	48
2. Visualización	53
2.1. Primera propuesta visual	54
2.2. Segunda propuesta visual	55
2.3. Propuesta final de la visualización	56
2.4. Mapa de relaciones en digital	57
3. Producto editorial en línea	58
VII. CONCLUSIONES Y RECOMENDACIONES	60
VIII. FUENTES DE INFORMACIÓN Y BIBLIOGRAFÍA	63
IX. ANEXOS	72
Figura 1. Primera propuesta de representación de películas venezolanas por género-temática	72
Figura 2. Segunda propuesta de representación de películas venezolanas por género-temática	73

Figura 3. Relación espacio-producto de las exportaciones de Venezuela (2012) - Observatorio de Complejidad Económica	74
Figura 4. Borrador final en papel del mapa de relaciones con 200 filmes por género-temática	75
Figura 5. Mapa de relaciones de películas venezolanas en formato digital	76
Figura 6. <i>Home</i> o portal del sitio web <i>Proyecto CIC-Big Data</i>	77

I. INTRODUCCIÓN

Desde que el término Big Data apareció hace casi un par de décadas, grandes volúmenes de datos empezaron a ser almacenados a diario y a una alta velocidad. Esto permitió que la información recopilada pudiera ser analizada y utilizada no solo para la toma de decisiones, sino también para otras actividades como determinar el rendimiento de estudiantes, realizar publicidad personal, diseñar nuevos materiales para la construcción, predecir epidemias y hacer pronósticos del clima. De esta manera, la data se había convertido en un elemento fundamental para el desarrollo de la sociedad.

Actualmente, los datos surgen de casi todas las actividades que el hombre realiza. El tomar una foto, grabar un video, hacer una transacción bancaria e, incluso, una llamada telefónica son solo algunos de los ejemplos de las cosas que producen grandes cantidades de información.

A pesar de la importancia que tiene la producción y recopilación de datos, no fue sino hasta finales del siglo XIX cuando comienza una verdadera preocupación por su almacenamiento. El largo proceso de análisis del censo realizado en Estados Unidos en el año 1880 y el crecimiento desmesurado de la demografía en ese país durante 1930 produjeron una sobrecarga de información que requería ser acumulada y organizada.

De igual forma, para poder responder al aumento en la demanda de investigaciones y publicaciones, las bibliotecas americanas necesitaban de nuevas y mejores formas de almacenaje. Es a partir de los años 40 cuando el *boom* o explosión de la información motivó a distintos científicos, como Claude Shannon, Fritz-Rudolf Güntsch, Derek Price, Howard Dresner y Edgar F. Codd, a plantear teorías y a generar métodos para guardar, utilizar, analizar y recuperar todos esos datos que se producían diariamente.

Esto conllevó a que durante los años siguientes se crearan bases de datos relacionales, memorias virtuales y tecnologías como el *Business Intelligence* y los sistemas de planificación de recursos de fabricación. Todo esto con el fin de buscar la mejor manera de guardar o conservar la información.

En la década de los 90, los computadores se caracterizaban por un gran tamaño y por una escasa capacidad de almacenaje, llegando a guardar un solo gigabyte de data. Con el paso del tiempo y con la evolución de la tecnología, los equipos fueron reduciendo sus dimensiones y aumentando sus velocidades de procesamiento. Esto, aunado al hecho de que dispositivos más económicos empezaron a archivar mayores volúmenes de información, fueron algunas de las razones que impulsaron el surgimiento del Big Data.

Aunque la producción y almacenamiento de datos se ha convertido en un proceso sencillo y cotidiano, la información generalmente está disponible solo para el consumo de las empresas. Sin embargo, esos datos podrían ser reutilizados para otros fines, como por ejemplo el periodístico. Por este motivo, ese material debería ser accesible para el público gratuitamente y sin ningún tipo de licencia o restricción. Esto es lo que se conoce como datos abiertos u Open Data.

Es relevante acotar que este último aspecto tiene que ver con que los datos sean utilizados, reutilizados y redistribuidos libremente por cualquier persona (Open Knowledge Foundation citado en Ferrer-Sapena y Sánchez-Pérez, 2013, p.150). Debido a esto, uno de los aportes del Big Data, y sobre todo del Open Data, es el uso de la información para generar trabajos para el Periodismo de datos.

Es en este ámbito en el que los datos cobran valor al ser usados por los periodistas para contar historias. De esta forma, reportajes, infografías e investigaciones se nutren de este material para revelar detalles, profundizar los hechos, establecer relaciones y mostrar qué hay más allá de la noticia.

Aunque en Venezuela el Periodismo de datos no es relativamente nuevo, sí es cierto que el limitado acceso a las fuentes oficiales y a datos nacionales e, incluso, internacionales es un reto al que se enfrentan los periodistas al momento de hacer trabajos de investigación.

Lo anterior es indicado por la periodista Emilia Díaz-Struck, quien agrega que si el periodismo venezolano se enfocara en el acceso a los datos y a la creación de bases propias, se formularían historias de interés general y se podría apoyar la investigación.

Debido al valor que representan las bases de datos para el desarrollo del Periodismo, sobre todo el de datos, la Universidad Católica Andrés Bello, específicamente su Centro de Investigación de la Comunicación (CIC-UCAB), decide emprender un proyecto, basado en bancos de data sobre cinematografía nacional, conocido como *Proyecto CIC-Big Data*.

El CIC-UCAB es un instituto enfocado en el estudio de diversas variables relacionadas con el fenómeno de la Comunicación Social en el siglo XXI. Por ello, siempre trata de ir de la mano con el empleo de nuevas tecnologías de la comunicación para fomentar una práctica periodística de calidad y de fácil comprensión.

Este trabajo de grado surge como resultado de una tesis-pasantía relativa al proyecto mencionado. Para este se toman en cuenta los conceptos de Big Data y Open Data. Sin embargo, ideas como la “musealización” y las “constelaciones de sentido” no pueden pasar desapercibidas, ya que son fundamentales para la concepción de una representación gráfica digital de los largometrajes que se han realizado en el país.

A pesar de lo anterior, ¿por qué realizar una visualización de filmes nacionales? Esta actividad es el punto central de la pasantía y pretende reflejar en forma de mapa de relaciones las películas según sus géneros y temáticas. De esta manera, el fin de este proyecto no solo es el de dar a conocer las piezas que componen la filmografía venezolana, sino además el de mostrar cómo estas obras de cine se relacionan de

acuerdo con los dos aspectos antes mencionados. Igualmente, otro de sus objetivos es el de compartir este contenido en la web para su uso, especialmente en el Periodismo de datos y en las futuras investigaciones que se realicen.

Es importante señalar que el presente trabajo escrito dedica siete de sus nueve apartados a describir el proyecto. El primero de ellos es la introducción, la cual se desarrolla en estas páginas. El segundo es el planteamiento del problema, en el que se presenta el *Proyecto CIC-Big Data* y su delimitación.

El tercer apartado es el marco contextual; allí se expone un panorama del Centro de Investigación de la Comunicación de la UCAB y su interés por realizar esta tesis-pasantía. De igual forma, se prosigue con el marco metodológico, en el que se plantean los objetivos, el método, diseño y tipo de la investigación, así como la breve descripción de las cuatro fases en las que esta se desarrolla.

El quinto punto consiste en el análisis y discusión de los resultados. En él se explica con detalle cómo es el proceso para la elaboración de la representación gráfica de las películas, lo cual incluye la construcción de categorías, de propuestas de visualización y de un producto editorial en línea. Finalmente, en el último apartado se encuentran las conclusiones y recomendaciones que surgieron al culminar el trabajo.

II. PLANTEAMIENTO DEL PROBLEMA

El Centro de Investigación de la Comunicación de la Universidad Católica Andrés Bello está desarrollando desde febrero de 2014 un proyecto digital enfocado en las tecnologías del Big Data y Open Data. Esto con el fin de compartir abiertamente un conjunto de información relativa a la cinematografía venezolana.

El CIC-UCAB en su intento por hacer accesibles estos datos decide poner en práctica el *Proyecto CIC-Big Data*, el cual contribuiría principalmente con el Periodismo de datos y con el de investigación. La idea de disponer de forma libre y gratuita en la web un repertorio de películas realizadas en Venezuela surge de la necesidad por la obtención de material fiable sobre este tema que pudiera servir de apoyo principalmente en la realización de trabajos de investigación.

Es importante mencionar que el centro inicia el proyecto con un enfoque en la cinematografía venezolana debido a que sobre este ámbito cuenta con una base de datos más elaborada y completa. Por este motivo, se quiere crear un mapa digital, diseñado con base en la musealización y en la teoría de las constelaciones de sentido de Walter Benjamin, que refleje no solo todos los largometrajes venezolanos, sino además la relación que tienen estos entre sí por géneros y temáticas.

Asimismo, la publicación de un portal en formato de blog serviría como plataforma para la difusión tanto de la base de datos, como del mencionado mapa. De igual forma, los dos productos contribuirían con las búsquedas, procesamientos y análisis de información de cine nacional.

Con relación a la pregunta de investigación, el presente trabajo de grado tiene como interrogante a ser respondida el saber ¿qué aspectos pueden desarrollarse del *Proyecto CIC-Big Data* a través de una tesis-pasantía?

También, de la pregunta anterior se derivan algunas otras interrogantes secundarias como son ¿qué es el Big Data y el Open Data y cómo a partir de estos se podría trabajar digitalmente con una base de datos de cine venezolano?, así como ¿cuál es la relación que se establece entre el Periodismo de datos y el de investigación con esta base?

La tesis-pasantía referente al *Proyecto CIC-Big Data* se desarrolla en la ciudad de Caracas entre los meses de octubre de 2014 y agosto de 2015, y está dirigida y supervisada por dos profesores del Centro de Investigación de la Comunicación de la Universidad Católica Andrés Bello. De igual manera, para el proyecto se trabaja con una base de datos que abarca 386 películas venezolanas producidas entre los años 1897 y 2015.

III. MARCO CONTEXTUAL

1. Centro de Investigación de la Comunicación de la Universidad Católica Andrés Bello (CIC-UCAB)

El Centro de Investigación de la Comunicación de la Universidad Católica Andrés Bello es una institución enfocada en el estudio de diversos fenómenos de la Comunicación Social contemporánea. Para este centro, la investigación es un recurso que se emplea para fomentar el saber y el conocimiento; por lo tanto, en él se realizan trabajos exhaustivos y de alto nivel. Fue fundado en el año 1992 y sus oficinas se encuentran en el piso cinco del Edificio Cincuentenario de la UCAB.

El CIC-UCAB pretende que sus investigaciones solventen las necesidades comunicacionales tanto de la universidad a la que presta servicio como las del país. Debido a lo anterior, promueve la continua actualización académica de sus colaboradores, presta asesoría a organismos comunicacionales en Venezuela y difunde los principales resultados de sus proyectos. En este sentido, uno de sus objetivos fundamentales es fomentar la práctica periodística de calidad.

El Centro de Investigación de la Comunicación cuenta con varios investigadores contratados, adscritos o asociados, los cuales son especialistas en distintas áreas de la comunicación. Entre ellos se encuentran Carlos Delgado Flores, actual director del centro, Andrés Cañizales, Humberto Valdivieso, Marcos Requena, Mariela Matos y José Pérez Quintero.

A lo largo de los años, el CIC-UCAB ha estado desarrollando diversos proyectos de investigación comprendidos en cinco áreas de estudio: Historia y memoria del Periodismo, Tecnologías de información y comunicación, Medios y representaciones sociales, Comunicación política y libertad de expresión, y Periodismo popular.

Durante los períodos académicos comprendidos entre 2013-2014 y 2014-2015, el centro ha realizado y promovido más de 50 proyectos vinculados a la Comunicación Social, los cuales están clasificados en siete nuevas áreas, variantes de las mencionadas anteriormente: Videoculturas, Comunicación organizacional integral, Imaginarios urbanos, Democracia y libertad de expresión, Estudios sociales y culturales de la comunicación, MediaLab UCAB y Sociedad del conocimiento y entorno digital.

En esta última área es donde se encuentra la tesis-pasantía para el desarrollo del *Proyecto CIC-Big Data*, cuyos marcos teórico y metodológico se desarrollarán en las siguientes páginas. Este trabajo es considerado por el CIC-UCAB como estratégico, ya que el centro no cuenta con la tecnología suficiente o la más adecuada para el almacenamiento de grandes cantidades de datos y su correcta representación gráfica.

Para este trabajo, el CIC-UCAB cuenta con la colaboración de la Escuela de Comunicación Social no solo para realizar una base de datos de películas venezolanas, desde sus inicios hasta las últimas que estén por estrenarse, sino también para la visualización gráfica de estas mediante el “cruce” de géneros y temáticas.

Con el acceso a todos estos datos, se mejoraría la calidad de las búsquedas relativas al tema y se precisarían los trabajos de investigación que se llegasen a realizar, promoviendo así el Periodismo de investigación y el de datos.

IV. MARCO TEÓRICO

1. *Big Data*

1.1. *Definición*

La producción de grandes conjuntos de datos está cada vez más ligada a las distintas actividades cotidianas que realizan las personas en la actualidad. El cúmulo de información que se genera en todos los contextos de la sociedad: aplicaciones de la web 2.0, los gobiernos, las investigaciones científicas, las publicaciones o las organizaciones crece de forma rápida y exorbitante y no puede ser manejado fácilmente por los sistemas informáticos (Ferrer-Sapena y Sánchez-Pérez, 2012, p.150). A esto se le conoce como Big Data.

Chris Sherman en *What's the big deal about Big Data?* (2014) lo definió como “un término que hace referencia a una colección de conjuntos de datos tan largos y complejos que resulta difícil procesarlos con herramientas tradicionales de bases de datos o con aplicaciones de procesamiento de datos” (p.2).

El Big Data o “datos masivos” forma parte del sector de las Tecnologías de la información y la comunicación, siendo allí uno de los campos más importantes. Estos datos son “grandes volúmenes de información que se mueven o analizan a alta velocidad y que pueden presentar una compleja variabilidad en cuanto a la estructura de su composición” (Tascón, 2013, p.48).

Soraya Paniagua (2013) explicó que los datos masivos son una “colección de tecnologías y estrategias capaces de capturar y analizar, de forma económica, grandes volúmenes de datos provenientes de múltiples fuentes heterogéneas a una alta velocidad” (p.96). Estos datos surgen ya sea de sensores ubicados en artículos de uso

cotidiano (o lo que se conoce como “Internet de las cosas”), las fotografías digitales, publicaciones en redes sociales, videos y compras en línea.

Como ejemplo de la gigantesca cantidad de data que se genera actualmente, se presentan cifras publicadas por McKinsey Global Institute (2011): en el año 2010 había cinco billones de teléfonos celulares en uso, 30 billones de piezas de contenido fueron compartidas en Facebook al mes y la Biblioteca del Congreso de Estados Unidos recolectó 235 terabytes (un terabyte es una unidad de almacenamiento de información y equivale a mil gigabytes) de datos.

Por esto, Francis X. Diebold (2000) describió al Big Data como una “explosión en la cantidad (y, a veces, la calidad) de los datos disponibles y potencialmente pertinentes, el resultado de los avances recientes y sin precedentes en el registro de datos y la tecnología de almacenamiento” (p.2).

En la actualidad, el Big Data ha pasado a representar una herramienta para el procesamiento de miles de datos que se generan diariamente. En la publicación *Una aproximación a Big Data* (2014), Javier Puyol Moreno señaló que en el mundo cada día se producen alrededor de 2.5 exabytes de datos, lo cual equivale a un millón de terabytes. Es decir, hay tanta información produciéndose que 90 % de todos los datos guardados hoy en día se produjeron en los últimos dos años.

Con relación al Big Data, Puyol Moreno indicó que este se aplica a “toda la información que no puede ser procesada o analizada utilizando herramientas o procesos tradicionales” (p.1). El autor estableció que el mayor desafío del Big Data se produce con relación a aquellos proyectos que estén vinculados a la toma de decisiones respecto de una gran cantidad de datos, en la definición de estrategias y en la obtención de mejores experiencias sobre el consumo de los individuos.

El Big Data es en resumen la mejor alternativa en el nuevo siglo para el procesamiento de los datos, ya que le permite a las empresas tomar las mejores decisiones en materia de análisis, mejoras en el mercadeo, y en la atracción de nuevos clientes.

1.2. Antecedentes

Una de las cuestiones que impulsó el surgimiento del Big Data fue la preocupación por el aumento constante y exorbitante de la producción de datos, debido al desarrollo de la actividad humana.

Para 1800, el lento procesamiento de los datos del censo de los Estados Unidos de América promovió la creación de la máquina tabuladora de Hollerith, la cual trabajaba con tarjetas perforadas y permitía analizar en un año estos datos que normalmente tardaban de ocho a 10 años en ser procesados.

De igual forma, el incremento exponencial durante el año 1932 de la demografía de este país, que conllevó a la emisión de números para el seguro social, así como al desarrollo del conocimiento, requerían mantener el registro de esa información de manera organizada. Es ante ello que surgió la necesidad de crear una nueva manera de condensar ese enorme volumen de datos, hacerlos manejables y comprensibles, para así analizarlos, “cruzarlos” y emplearlos en el estudio de los fenómenos sociales.

En el año 1941, apareció por primera vez el término “explosión de la información”, en un artículo del periódico Lawton Constitution. Es a partir de ese momento que se reconoció la dificultad que existía para poder manejar ingentes cantidades de data.

Debido a lo anterior, se comenzaron a desarrollar nuevas técnicas que permitieron recopilar y gestionar grandes volúmenes de contenido. Asimismo, surgieron

teorías, pioneros, ideas y avances tecnológicos que impulsaron lo que hoy se conoce como Big Data.

Entre estos estaba la teoría de la información de Shannon, la cual determinaba cuáles eran los datos mínimos para transmitir información a través de canales afectados por ruido. Con esto se logró reducir el volumen de data, la cual sería mucho mayor que la que se conoce actualmente.

Para el año 1956, el físico alemán Fritz-Rudolf Güntsch desarrolló la memoria virtual, la cual permitía el almacenamiento de mayores cantidades de datos sin las limitantes que provocaba la memoria de hardware.

A principios de los años 60, el experto en información Derek Price consideró que los volúmenes de conocimiento científico estaban duplicándose cada 15 años e incrementándose por 10 cada 50 años.

Así como hubo un aumento de la información en el área científica, para 1966 también lo hubo en los negocios, motivo por el cual muchas empresas decidieron invertir en el diseño y desarrollo de sistemas informáticos que les ayudaran con los registros de inventarios.

Llegada la década de los 70, Edgar F. Codd, un matemático de Oxford que laboraba en IBM Research Lab, postuló la base de datos relacional. Esta propuesta indicaba la manera en la que se podía acceder a información almacenada en bases de gran tamaño sin tener conocimiento de cómo estos datos estaban estructurados o en dónde residían.

Para ese momento el acceso y recuperación de información era una tarea ardua y exigente de tiempo, que requería de especialistas con conocimientos en informática.

Sin embargo, para 1983 se produjeron avances en los sistemas tecnológicos, lo que benefició a todos los sectores, ya que en ese momento estos poseían nuevas formas de organización, almacenaje y generación de data. De esta manera, las empresas podían sacar mayor provecho a la hora de tomar decisiones.

Dos años más tarde, los investigadores de IBM Barry Devlin y Paul Murphy construyeron una arquitectura para informes y análisis de negocio para la empresa. Esto fue lo que se convirtió en la base del almacén de datos o *data warehouse*.

Casi entrando a los 90, se retomó el término *Business Intelligence* o inteligencia empresarial, acuñado en 1958 por Hans Peter Luhn. Fue ampliado por Howard Dresner, quien lo definió como aquellos conceptos y métodos que mejoran la toma de decisiones relativas a negocios con el uso de sistemas de apoyo que se basan en datos reales.

En 1992, surgió el primer informe creado por Crystal Reports sobre bases de datos. Esto permitió a las empresas hacer uso del *Business Intelligence* que, de acuerdo con Mario Tascón (2013), “es el conjunto de estrategias y herramientas que una empresa tenía a su disposición para poder analizar los datos de su organización” (p.48).

Para la década de 1990, hubo un auge en el crecimiento tecnológico, por lo que los datos de la inteligencia empresarial se empezaron a condensar en documentos de Microsoft Excel. Sin embargo, algunas empresas no consideraban adecuados los sistemas que utilizaban para apilar la información, por lo que comenzaron a crearse softwares y aplicaciones para gestionar esos datos.

La evolución tecnológica permitió la evolución de lo que hoy es el Big Data, el cual, según Tascón, está asociado a la minería de datos: “Un campo de las Ciencias de la Computación que intenta descubrir patrones en grandes volúmenes de datos” (p.48). De acuerdo con este autor, esta forma parte del *Business Intelligence* y utiliza métodos de la inteligencia artificial y la estadística para analizar los patrones en las bases de datos con las que trabaja (p.48).

La primera vez que se empleó el término Big Data fue en julio de 1997. Este apareció en un artículo redactado por los investigadores de la NASA Michael Cox y David Ellsworth, quienes indicaban que el ritmo con el que crecían los datos en ese momento se estaba convirtiendo en un problema para los sistemas informáticos de la época; a esto lo nombraron “el problema del Big Data”. (Cox y Ellsworth, 1997, p.1).

Asimismo, otro problema en la gestión de la data fue el alto coste que implicaba su publicación y almacenaje. Sin embargo, el almacenamiento en digital resultaba mucho más económico, lo que produjo el surgimiento de la plataforma de inteligencia empresarial. Esto fue reforzado por Michael Lesk, quien indicó que para el nuevo milenio no sería necesaria la eliminación de información, ya que la producción de cinta y disco alcanzaría un volumen de miles de petabytes para guardarla.

En el año 2000 se crearon las empresas de software para el desarrollo y gestión de bases de datos como son Oracle y SAP. Finalmente, en el año 2006 surgió Hadoop, la cual sirve como herramienta para procesar la explosión de datos que empezó a mediados de la primera década del siglo XXI.

1.3. Características

El Big Data posee características que le permiten funcionar adecuadamente y almacenar grandes cantidades de datos. Entre sus principales cualidades se encuentran las tres “v” o volumen, velocidad y variedad.

El volumen de almacenamiento de datos es uno de los principales rasgos del Big Data. Con cada día que pasa se sigue expandiendo y el lenguaje utilizado para referirse a este cambia.

De acuerdo con Eaton et al (2012) en *Understanding Big Data*, “la conversación acerca de volúmenes ha cambiado de terabytes a petabytes con un inevitable cambio a zettabytes, y toda esta data no puede ser almacenada de manera tradicional” (p.7).

El hecho de que cada día se creen 2,5 quintillones de bytes refleja la continua expansión o “explosión” del volumen del Big Data. Asimismo, lo es el que 40 zettabytes de datos serán generados para el año 2020; es decir, habrá un aumento de 300 veces la data que existía en el 2005 (IBM).

Con relación a la variedad, segunda cualidad del Big Data, Corrigan et al (2013) explicaron que esta “trata de capturar todos los datos que pertenecen a nuestro proceso de toma de decisión” (p.9).

Es decir, la variedad representa todos los tipos de data. “Un cambio fundamental en los análisis de requerimientos de las estructuras tradicionales para incluir datos en bruto, semiestructurados, y no estructurados como parte de la toma de decisión y el proceso de discernimiento” (Eaton et al, 2012, p.8).

Los diferentes tipos de data, los cuales forman parte de esta variedad, pueden provenir de sitios web, correos electrónicos, búsquedas en Internet, documentos, sensores, piezas de contenido compartidas en Facebook, videos y mensajes en Twitter.

Por último, la velocidad se define como el ritmo con el que se mueve o analiza la data (Tascón, 2013, p.48). De igual manera, Corrigan et al (2013) coincidieron con que esta se refiere al “ritmo con el que los datos llegan a la empresa y son procesados o bien comprendidos” (p.10).

Sin embargo, una cualidad que no puede pasar desapercibida es la veracidad, la cual es incluida por algunos autores como una cuarta “v” del Big Data. Corrigan et al explicaron que este término se emplea cada vez más para describir los datos masivos, por lo que se utiliza para referirse a la calidad o integridad de estos (p.14).

Según la publicación *Analytics: el uso de Big Data en el mundo real* del IBM, la veracidad tiene que ver con la incertidumbre de la información:

“La veracidad hace referencia al nivel de fiabilidad asociado a ciertos tipos de datos. Esforzarse por conseguir unos datos de alta calidad es un requisito importante y un reto fundamental de Big Data, pero incluso los mejores métodos de limpieza de datos no pueden eliminar la imprevisibilidad inherente de algunos datos, como el tiempo, la economía o las futuras decisiones de compra de un cliente” (p.5).

El disponer de unos datos certeros y correctos es de suma importancia para las empresas, ya que de acuerdo con IBM, uno de cada tres líderes no confía en la información que emplea para tomar decisiones, además, 27 % de los encuestados no estaba seguro de cuán errada era la data.

Aunque el autor Mario Tascón indicó que las características mencionadas anteriormente definen lo que es el Big Data, él incorpora una última que es la visualización. Esta “no solo forma también parte de ello, sino que muchas de las imágenes que nos traen a la memoria el trabajo con Big Data tienen que ver con estas nuevas formas de ‘ver’ estos datos” (p.48).

Debido a la excesiva cantidad de datos que se producen diariamente, su representación o visualización es una herramienta para que las empresas puedan comprender y analizar toda la información, especialmente cuando deben tomar decisiones sobre negocios.

1.4. Tipos de datos

La fuente fundamental del Big Data son los datos. Mario Tascón señaló que estos son “el petróleo del siglo XXI”, por su importancia y valor tanto para las empresas como para la sociedad en general. Dentro de estos tipos de datos se encuentran los estructurados y los no estructurados:

Los datos estructurados son, según este autor, todos aquellos “que provienen de fuentes de información conocidas y que, por tanto, son fáciles de medir y analizar a través de los sistemas tradicionales” (p.48). Es decir, en esta categoría se incluyen las bases de datos.

Chris Sherman (2014) agregó que los datos estructurados pueden ser ingresados en bases de datos u hojas de cálculo que tienen previamente definidas celdas y campos. Asimismo, estos “pueden ser (...) manipulables y analizados usando fórmulas y algoritmos diseñados para encontrar las mejores respuestas” (p.11).

Los no estructurados, por el contrario, son “aquellos datos no almacenados en una base de datos tradicional”, de acuerdo con Juan Vidal en el sitio web DataPrix. Para él, la información no estructurada no puede ser almacenada en estructuras de datos relacionales predefinidas.

De igual forma, tanto Sherman como Tascón acotaron que los no estructurados consisten en textos, números, imágenes y videos con los que la web está hecha, así como los datos que llegan de la Web, de las redes sociales, sensores de las ciudades y edificios.

Un ejemplo desde el punto de vista médico que ilustra todo lo anterior es el expuesto por Soraya Paniagua (2012):

“Un dato estructurado es un dato que puede ser almacenado, consultado, analizado y manipulado por máquinas. Un dato desestructurado es todo lo contrario. Por ejemplo datos no estructurados son las recetas de papel, los registros médicos, las notas manuscritas de médicos y enfermeras, las grabaciones de voz, las radiografías, resonancias magnéticas, TAC y otras imágenes. Los datos estructurados y semiestructurados incluyen archivos electrónicos de contabilidad, datos de actuario o datos clínicos”.

Además de todo lo anterior, hay que mencionar que existe una categoría intermedia entre estos tipos de datos que es la de los semiestructurados. Estos son los que poseen una estructura que está implícita, pero que no es tan regular como para ser gestionados como estructurados. Dentro de estos se encuentran las páginas web que trabajan con HTML y la señalética en general.

A manera de resumen, los datos estructurados serían aquellos fácilmente procesables y analizables con herramientas tradicionales y los no estructurados, aquellos que por su forma no pueden registrarse en estas.

1.5. Big Data en América Latina

El desarrollo de la actividad humana ha dejado a su paso un gran cúmulo de información. La contabilización de los censos nacionales, los resultados en investigaciones científicas, el número del seguro social de los ciudadanos, la cantidad de votos obtenidos en unas elecciones, los libros disponibles en una biblioteca universitaria, y hasta los datos personales que manejan los grandes bancos generan un volumen considerable de data.

Hace unos años, una computadora y una hoja de cálculo bastaban para manipular las cantidades de información que se generaban. Sin embargo, ahora es necesario manejar los enormes conjuntos de datos desde la “nube” o *Cloud Computing*, es decir, herramientas de almacenamiento usadas a través de Internet. La data ya no se almacenaría en algún ordenador físico, sino que estos datos serían operados desde servidores en Internet.

La información está en todos lados y según consideró Peter Kroll (citado en Pérez Arbesú, 2013), director de la firma Capgemini en México, esta es tan valiosa que es considerada como una nueva mina de oro; de hecho, afirma que Big Data ahora es sinónimo de *Big Business*.

De acuerdo con Jaime García Cantero (2013), la edición de mayo de 2011 de la revista *Science* calculó que “la información digital almacenada en el año 2007 en todo el mundo era de más de 276 mil millones de gigabytes, unas 13.800 veces la información almacenada 20 años antes” (p.52). También, señaló que el contenido generado en texto, video e imágenes ha crecido de forma exponencial.

Este autor afirmó que “se necesitarían 12 años seguidos sin parar a dormir ni a comer para ver todos los vídeos que se subieron ayer a Youtube, y a Facebook se suben más de tres mil fotografías cada segundo” (p.52).

A pesar de la importancia de esta nueva fuente de dinero e inversión, muy pocas empresas en América Latina se atreven a incursionar en el uso de los datos masivos. De hecho, una encuesta realizada por TechTarget sobre las prioridades de las Tecnologías de la Información en 2013 mostró que la región se encuentra rezagada en el uso de Big Data y la analítica.

Los resultados publicados por la encuesta arrojaron que 32 % de las empresas consultadas todavía no tenían planes de incluir este tipo de tecnología, y 23.9 % apenas estaba evaluando la posibilidad de implementar, emplear y manejar grandes datos. Únicamente, 36,4 % de las empresas que respondieron empezarían a usar los datos masivos, los expandirían o seguirían trabajando con los programas que poseen actualmente.

Esta situación se presenta a menudo por confusiones en el significado de Big Data. Esto lo aseguró Lizzette Pérez Arbesú (2013), directora de TechTarget en la región, quien además indicó que se suele pensar en el término como el manejo de grandes sumas de datos generados por parte de enormes empresas, cuando en la realidad la idea va más allá. Por otro lado, Kroll dijo que en el caso de los clientes mexicanos, estos no tienen datos confiables o no cuentan con bases sólidas, por lo que “la inclusión de Big Data solamente les generaría más confusión” (citado en Pérez Arbesú, 2013).

Otra encuesta de TechTarget para 2015 arrojó que el Big Data se estaba convirtiendo en la cuarta prioridad de las Tecnologías de la información para 27,6 % de las empresas encuestadas. Es decir, casi 10 % más compañías empezarían a implementar estos datos en comparación con 18 % para el año 2014.

Es importante resaltar que las inversiones en esta materia no se han dejado de realizar. César Longa, gerente de Programa de Software para IDC (International Data Corporation) Latinoamérica, estimó que la inversión de Big Data representó alrededor de 13 mil millones de dólares, siendo México y Brasil los países que lideraron la región con 820 millones de dólares por cuestiones de almacenamiento de datos durante el 2014. Asimismo, el representante pronosticó que dentro de cinco años el mercado de la data se expandirá en 232 % en términos regionales (citado en *Mundo Contact*, 2014).

Algunos países de la región latinoamericana han tomado la decisión de sumarse al uso de estas nuevas tecnologías para mejorar la productividad y el desarrollo social; sin embargo, es necesaria una mayor explotación de estas herramientas para lograr y promover proyectos, obtener información de utilidad y así emplearla para tomar las mejores decisiones.

2. *Open Data*

2.1. *Definición*

La ingente cantidad de información que se produce y que se duplica rápidamente perdería su valiosa utilidad si no se permitiera el libre acceso a los datos que la componen. “Ya no se trata solo de compartir información, ahora (se quiere) que sean accesibles también los datos, de manera libre y gratuita”, afirmaron Ferrer-Sapena y Sánchez-Pérez (2012, p.150).

Por ello, se pretende la transparencia en la producción de estos datos, independientemente de la fuente de la que vengan (empresas, Estados o la Academia). Se quiere que sean fáciles de entender, reusables y que estén a la disposición del mayor número de personas.

De ahí que surgiera el término Open Data, definido por los autores antes citados como “la apertura (o puesta a disposición de cualquiera) de datos, sin restricciones de *copyright*” (p.150). Es decir, sin mayores limitaciones legales o de derechos de propiedad.

La Open Knowledge Foundation (2012) explicó que este término se refiere a “todos los datos que pueden ser utilizados, reutilizados y redistribuidos libremente por cualquier persona y que se encuentran sujetos, cuando más, al requerimiento de atribución y de compartirse de la misma manera en que aparecen” (p.6). Así, la gente pueda darle el uso que desea, pero otorgando el debido reconocimiento al autor

Esta fundación también aclaró que la apertura de esta data se hace exceptuando los datos personales: “No contiene información sobre individuos específicos” (p.6).

Tapia (s.f.) sostuvo que “los datos deben publicarse en bruto (sin procesar), bien estructurados y en formatos conocidos que faciliten la reutilización”. Es por eso que no

basta con las publicaciones de libre acceso o las informaciones que normalmente se consiguen en un sitio web (p. 6). De igual forma, esta apertura de datos permite que los usuarios puedan crear nuevos contenidos a partir de ellos, “que (resulten) en nuevos datos, conocimientos, mejorar procesos, dar valor añadido a los existentes” (Tapia, s.f., p.6).

Marc Garriga-Portolà (2011) recalcó que lo que se está proporcionando no son simples documentos en formato Pdf o Microsoft Word, sino datos propiamente dichos. “Un servicio Open Data no es un servicio *open access*, sino un sitio donde se ofrecen datos en bruto o en crudo (*raw data*) para poder ser reutilizados y crear otros servicios públicos” (p.301).

Este experto agregó que es indispensable presentar los datos de forma adecuada para que puedan ser leídos por los motores de búsqueda y así las personas tengan la capacidad de encontrarlos y darles el uso que deseen. Mariño (2012) señaló que estos datos abiertos sin procesar “pueden tratarse de informes, de bases de datos, de encuestas, de estadísticas y demás, en formatos compatibles con los potenciales usuarios y terceros”.

Asimismo, Ferrer-Sapena y Sánchez-Pérez (2012) indicaron que estos datos son en su mayoría “geográficos, estadísticos, genómicos, médicos y biológicos” (p.150). Sin embargo, estos autores no centraron su atención en los tipos de datos que se abren, sino en que estos estén disponibles en formatos adecuados para realizarles transformaciones que sean útiles para los usuarios.

2.1. Características

Entre las cualidades que definen lo que es Open Data se encuentran “la disponibilidad y el acceso”, es decir, que los datos estén al alcance de las personas mediante el uso de Internet, siempre que su formato pueda cambiarse (Ferrer-Sapena y Sánchez-Pérez, 2012, p.151).

Igualmente, estos autores presentaron como elementos fundamentales en el proceso de abrir los datos “la reutilización y redistribución”, lo que significa que esta información debe poder “cruzarse” con otra para formar combinaciones, producir informaciones nuevas y de mayor valor, y así entender y explicar un fenómeno de un modo más completo (p.151).

Ferrer-Sapena y Sánchez-Pérez (2012) destacaron el riesgo de que esta conjunción de datos se use para fines que no fueron los pensados originalmente y termine siendo utilizada por terceros para el detrimento de otros. A pesar de esto, reconocen que “el movimiento *Open Data* exige que los datos se proporcionen sin ningún tipo de restricción, exigiendo la no discriminación en la concesión de licencias”, sin importar qué persona u organismo los vayan a emplear (p.153).

Garriga-Portolà (2011) señaló que las limitaciones de reutilización de los datos solo deben darse “por razones de seguridad, de privacidad o (porque estén) regulados por una ley específica o por un procedimiento administrativo” (p.300).

Como última categoría se presenta “la participación universal”, es decir, que toda persona pueda tener la posibilidad de usar estos datos, sin tener que pagar por el acceso a ellos (Tapia, s.f., p.6).

La Open Knowledge Foundation (2012) afirmó al respecto que “no debe haber discriminación contra (...) personas o grupos”. Esta fundación citó como ejemplo las restricciones de tipo no comercial, que prohíben obtener ganancias económicas del uso de los datos, y aclaró que dichas condiciones no tienen cabida dentro de la definición de datos abiertos (p.6).

Todas estas características son lo que permite que la apertura de datos sea un proceso útil y necesario para el usuario que está ávido de contenidos significativos, de transparencia en la información y de mejorar su experiencia en la web.

3. Periodismo de datos

3.1. Definición

En el *Manual de Periodismo de Datos*, Paul Bradshaw, de la Birmingham City University, definió este como aquel que tiene el fin de ayudar al comunicador social, ya sea a contar una historia convincente mediante infografías que resulten atractivas, a explicar la relación entre una historia y un determinado individuo, a abrir procesos de búsqueda de información, así como a compartir preguntas y datos en los blogs.

Asimismo, Bradshaw explicó que los datos pueden ser la fuente de este tipo de Periodismo y pueden ser la herramienta con la que se narra la historia o, incluso, ambas cosas. “Como cualquier fuente, debe tratarse con escepticismo; y como cualquier herramienta, debemos ser conscientes de cómo puede modelar y limitar las historias que se crean con la misma”.

Según Sandra Crucianelli (2013), el Periodismo de datos es una manera de hacer Periodismo a partir del uso de nuevas tecnologías y de la gran cantidad de información y de datos que circulan por la autopista informática, es decir por la web.

La periodista de investigación y precisión señaló que este tipo de Periodismo se caracteriza no solo por la interactividad –ya que es el usuario el que decide si descarga o no un contenido o los datos disponibles en Internet–; sino, además, por estar compuesto por otras disciplinas como son el Periodismo de investigación, de profundidad, de precisión, el analítico y el asistido por computadora (2013, p.107).

Por este motivo, el Periodismo de datos es a la vez de investigación porque se usan técnicas propias de este. Tiene algo del de profundidad, dependiendo del tema que se analice y es de precisión, ya que a veces se debe recurrir a los métodos de investigación social (2013, p.107).

También, es asistido por computadora debido a que se deben emplear hojas de cálculo, datos y procedimientos propios de esta categoría. Finalmente, es Periodismo analítico, ya que muchas veces se deben tomar los datos y llevarlos a sistemas de información geográficos que permitan medir mejor el comportamiento de las variables que se manejan (2013, p.108).

El Periodismo de datos, junto con el surgimiento de tecnologías de información y computación, ha permitido analizar grandes volúmenes de datos, con lo cual se pueden influenciar los procesos de toma de decisiones y la creación de políticas públicas para desarrollar una sociedad del conocimiento (Diario *La Nación*, 2012).

“Las herramientas tecnológicas hoy abren grandes posibilidades para que los periodistas, y en particular los periodistas de investigación, aporten un valor diferencial y lleven al Periodismo a un nuevo estadio en su evolución”, estableció Florencia Bianco, gerente de comunicaciones de Google Argentina, ante el hecho de que el Periodismo de datos y los periodistas pudieran ser desplazados por la llegada de los avances en tecnología (citada en Diario *La Nación*, 2012).

Tomando en consideración lo anterior, el Periodismo de datos se puede entender como la nueva manera de hacer Periodismo caracterizada por el empleo de grandes cantidades de datos y cifras analizadas con softwares y aplicaciones tecnológicas con el fin de contar nuevas historias haciéndolas atractivas y comprensibles para el público.

3.2. Periodismo de datos como herramienta para la investigación

Con cada día que pasa, los avances en materia de aplicaciones tecnológicas son mayores. Con esto, el Periodismo de datos también se desarrolla y evoluciona para ofrecer mejores servicios, trabajos y datos de calidad a los distintos lectores e investigadores.

Este se ha convertido en un elemento indispensable para el manejo de grandes cantidades de datos en la web. Las fotos, comentarios, opiniones y números forman parte del conjunto de elementos que pueden ser manejados por esta práctica para presentar noticias de manera distinta y, al mismo tiempo, para que los usuarios puedan tener a la mano cualquier dato que necesiten para corroborar o ampliar una información.

El tipo de Periodismo que se ha desarrollado permite la creación de nuevos enfoques para narrar historias. Aron Pilhofer del diario *New York Times* consideró que el Periodismo de datos es un término que incluye un conjunto de herramientas, técnicas y enfoques de la narrativa:

“Puede incluir todo, desde el tradicional Periodismo asistido por computadoras (usando datos como una “fuente”) hasta la visualización más avanzada de datos y aplicaciones de noticias. El objetivo unificador es periodístico: Proveer información y análisis para ayudar a informarnos todos sobre asuntos importantes de actualidad”.

De acuerdo con Jerry Vermanen, dataperiodista del sitio web Nu.nl, el Periodismo de datos implica “un nuevo conjunto de habilidades para buscar, comprender y visualizar fuentes digitales, en una época en que las capacidades básicas del Periodismo tradicional ya no bastan. No lo reemplaza, le agrega cosas”.

Asimismo, estableció que el periodista no solo debe mantener mayor contacto con las fuentes, sobre todo en este momento en el que estas se están volviendo digitales, sino que siempre deber saber cómo encontrar, analizar y visualizar una historia a partir de datos:

“El Periodismo de datos es solo el comienzo de la evolución de nuestras prácticas pasadas para adaptarse al *online*. El Periodismo de datos sirve a dos importantes propósitos para las organizaciones de noticiosas: encontrar historias únicas (no de los cables) y ejecutar la función de alerta. Especialmente en tiempos de crisis financieras, estos objetivos son importantes para los diarios”.

De igual forma, Nicolas Kayser-Bril, de la agencia Journalism++, expresó que este Periodismo es una herramienta que ha llevado a quienes se dedican a tomar las decisiones a cuantificar los avances de sus políticas, a monitorear las tendencias y, también, a identificar las oportunidades para la investigación.

Además, expresó que el manejo de datos por parte de los periodistas les ayudará a tener un sentido agudo al momento de manejar cifras, e inclusive podría beneficiarles para interactuar con el departamento de Relaciones Públicas.

Jesús Flores Vivar citado en Tapia indicó que el Periodismo de datos consiste en usar herramientas estadísticas y de visualización para profundizar las viejas historias y descubrir nuevas historias que contar (p.12). Ampliando esta idea, Mirko Lorenz, de Deutsche Welle, dijo que los periodistas deberían ver los datos como una oportunidad.

Con estos datos se pueden analizar la dinámica de situaciones complejas, tales como un disturbio o un debate político, o se podría observar cómo un fenómeno afecta a la gente de acuerdo a su edad, su género o su nivel de educación. “Usar datos transforma algo abstracto en algo que todos pueden entender y con lo que pueden relacionarse”, comentó el periodista.

De esta forma, el empleo de los datos es de relevancia para la investigación, puesto que permitiría que los periodistas analicen con más detalle situaciones diversas, ofreciendo así nuevos enfoques.

3.3. Periodismo de datos y su uso en Venezuela

Acceder a los datos que se necesitan para desarrollar un trabajo periodístico de calidad, como los provenientes de fuentes oficiales y de organizaciones privadas, es cada vez más difícil. Aunque estos son los pilares de muchos artículos de investigación, en especial de los de índole periodística, están velados u “ocultos bajo distintas capas de información” (Crucianelli, 2013, p.108). Por lo que el periodista es el principal

encargado de la búsqueda, recopilación y traducción de esos datos para la comprensión del lector.

Crucianelli sostuvo que el hecho de ocultar los datos se observa sobre todo en “países en los que no existen leyes de acceso a la información pública” y que son las naciones hispanohablantes las que principalmente no están acostumbradas a hacer públicos sus datos oficiales para facilitárselos a los ciudadanos (2013, p. 108).

La periodista especializada en Periodismo de investigación señaló que los datos “son una condición necesaria sin la cual no se puede siquiera comenzar a diseñar hipótesis periodística alguna” (p. 109). Allí radica la importancia capital de la obtención y el manejo de datos.

Venezuela se encuentra entre los países que no tienen una ley de acceso a la información pública. Emilia Díaz-Struck (s.f.) en el *Manual de Periodismo Iberoamericano* indicó que “falta generar una cultura que promueva la transparencia y el uso de datos abiertos disponibles vía electrónica”. La autora agregó que se puede buscar información en las gacetas oficiales, las sentencias jurídicas y las empresas a las que contrata el Estado, pero “los formatos no siempre son los ideales”. Sin embargo, esos datos “sirven de punto de partida para el trabajo periodístico”.

Además, la periodista Díaz-Struck afirmó que la diversidad de datos e información aumenta cuando los casos de investigación venezolana están vinculados a otros países, debido a las leyes de acceso a la información. “El abordaje que se ha dado en Venezuela, a partir del contexto actual, está vinculado con las posibilidades que existen al conectar historias con otros países, así como al generar tablas de datos propias a partir de documentos”.

El Periodismo de datos contribuye a desarrollar y presentar trabajos de investigación sólidos que ayudan a desmontar rumores. Por lo tanto, fomenta la estabilidad en la población. En el caso venezolano, Díaz-Struck destacó como pioneros

del empleo de este tipo de Periodismo los trabajos que involucraron el uso de redes sociales, la visualización de datos, los reportajes y las alianzas.

Ejemplos de estos son el portal web que creó el diario *El Nacional* con motivo de las elecciones presidenciales del año 2012, el cual llamó *Tweetómetro* y con el que realizó un seguimiento al impacto que tenían los candidatos Hugo Chávez y Enrique Capriles Radonski en la red social *Twitter*.

Asimismo, está el reportaje *Leche importada por el Gobierno se va de contrabando a Colombia* del portal web *El Mundo: Economía y Negocios*, el cual utilizó bases de datos de Venezuela y Nueva Zelanda (el país de donde provenía la leche importada) para corroborar “la diferencia reportada por el Instituto Nacional de Estadística, las bases de datos de movimientos portuarios de Puerto Cabello, Importgenius y Statistics New Zealand”, lo que puso de manifiesto el caso de contrabando (Díaz-Struck, s.f.).

El trabajo periodístico *Acuerdo crudo por derivados entre Ecuador y Venezuela benefició a intermediarios* publicado por el diario ecuatoriano *El Universo* fue una alianza entre estos países suramericanos, y demostró que dicho acuerdo no impidió que las empresas intermediarias aumentaran los precios de venta.

Para este artículo se empleó “una tabla de datos en la que se vació la información de las facturas de PDVSA, así como los registros portuarios y certificados de origen de los productos que llegaron a Ecuador” (Díaz-Struck, s.f.).

El Periodismo de datos es una disciplina que se está desarrollando poco a poco en Venezuela, pero que cada vez más está alertando a los periodistas sobre su utilidad social y en los procesos de investigación.

3.4. *Big Data y su aporte al Periodismo de datos*

La herramienta de almacenamiento Big Data, que permite manejar grandes valores de información, promueve la investigación, sobre todo interpretativa, así como el Periodismo en general, ya que facilita tener al alcance gran cantidad de datos e informaciones sociales y económicas de manera rápida y efectiva.

Esto es de mucha importancia para grandes empresas, centros de investigación y para cualquier persona que esté interesada en el uso de cifras y contenidos para la elaboración de un trabajo, debido a que muestra de manera dinámica e interactiva contenidos que pueden “cruzarse” entre sí para facilitar la comprensión de fenómenos sociales surgidos en el mundo.

La profesora Alicia Tapia sostuvo que “lo verdaderamente valioso [de un dato] es contextualizarlo y situarlo, saber qué significa y de ahí poder construir una historia con fundamento” (p.12). El orden es lo que da sentido a los datos, permite interpretarlos, traducirlos de manera correcta para formar un trabajo de investigación coherente y de calidad. “Los datos unidos y cruzados pueden desvelar realidades escondidas hasta entonces. Eso es lo que hace el periodismo de datos y la visualización de datos”, indicó Mar Abad (2012).

Por ello, Ángeles Mariño (2012) señaló que la aparición del *Big Data* supuso cambios en la forma de realizar Periodismo y en el modo de contar las historias, y que los datos de manera aislada no aportan mucha información, pero que “a nivel masivo sí [son importantes] cuando se los analiza desde el ángulo correcto”.

Es decir, la relevancia de los datos radica en el hecho de que estén agrupados, ya que así se podrían realizar interrelaciones, “cruces” o análisis para comprender fenómenos, cosa que sería más difícil al manejar datos de manera individual, puesto que estos por sí mismos no aportan información completa sobre un hecho cualquiera.

4. *Walter Benjamin*

Walter Benjamin (1892-1940) fue un filósofo y crítico social y literario alemán, cuya obra ha permitido entender los procesos que definen y describen “la contemporaneidad: La mercantilización generalizada, las nuevas formas cognoscitivas, la crisis de la experiencia histórica tradicional y las propuestas estéticas en un contexto tecnológico avanzado” (Círculo de Bellas Artes, 2011).

Aunque Benjamin no desarrolló una teoría del conocimiento propiamente dicha, “las sugerencias epistemológicas que aparecen dispersas en sus escritos” son el pilar en el que se apoya su pensamiento crítico (Círculo de Bellas Artes, 2010).

“La yuxtaposición dialéctica de elementos autónomos –imágenes, textos, evocaciones, recuerdos– genera una constelación de sentido”, es decir, “una retícula de conexiones significativas entre elementos independientes y distantes”. Con este proceso, el filósofo consiguió establecer relaciones entre cosas diferentes, dándoles así un nuevo valor. Benjamin llamó a este proceso del conocimiento “una iluminación profana” (Ibídem).

César Rendueles y Ana Useros (2010) explicaron que este crítico literario alemán empleaba un método parecido al elaborar su trabajo, que se basaba en “un complejo sistema de referencia, archivado e indexado, inspirado en el Atlas Mnemosyne del historiador Aby Warburg”. Benjamin se valía de “señales y diagramas” para conectar sus propias obras, citas de otros autores o distintos conceptos, creando así una constelación de sentido (p.12).

Este método de Benjamin “buscaba lo universal en el objeto material”. Para lograrlo, identificaba semejanzas en “las cosas singulares” que luego “se le mostraban al observador en constelaciones” (John, 2010, p. 81).

Horacio Delbueno (2010) señaló que la teoría de las constelaciones está formada por “elementos rescatados del pasado, resignificados [conectados] con los elementos del presente (...), tal cual manifiesta el propio Benjamin en su ensayo *La dialéctica en suspenso*” (p.38).

El intelectual alemán sacaba a la luz constelaciones teóricas insospechadas, mediante el engranaje conceptual de ideas, textos o materiales aparentemente alejados entre sí (Círculo de Bellas Artes, s.f.). Lo que daba como resultado un entretejido de conexiones, a modo de red neuronal, que ponía de manifiesto las relaciones existentes entre los distintos puntos que formaban la constelación.

Sin embargo, Fredric Jameson (2011) sostuvo que este tipo de referencias cruzadas y simultáneas en Benjamin no se usan para unir distintos conceptos, sino para establecer diferencias entre ellos. Agregó que cada punto en la constelación debe verse “como un modo de relación, en la que una parte permanece en su lugar como parte”, y a la vez esta le da forma al todo que es la constelación.

La combinación de las distintas partes que componen la constelación es lo que define la totalidad de la red neuronal y lo que le da un nuevo significado a cada uno de los puntos. (Manchado, 2014). Además, los conceptos pueden cambiar de valor y significado, al cambiar de posición dentro de la constelación. (Rendueles y Useros, 2010, p.12).

Debido a todo lo anterior, las constelaciones de sentido de Walter Benjamin se pueden considerar como un modelo que logra conjugar y conectar diversas categorías con variables distintas, ya que las redes neuronales permiten cruzar la mayor cantidad de estas para mostrar las diferentes interconexiones existentes.

5. *Mindomo*

Mindomo, cuyo sitio web es www.mindomo.com, corresponde a una pequeña compañía de software colaborativo en línea. Fue creada en el año 2007 y su principal objetivo es el de servir de herramienta para crear y compartir mapas mentales.

Esta aplicación, que está disponible en Windows y en los sistemas operativos iOS, Android y Linux, es útil, dinámica y sencilla. Permite al usuario realizar presentaciones, trabajar con colaboradores y crear mapas a los que se les puede insertar notas, íconos, elementos multimedia, hiperenlaces y elementos adjuntos.

El software cuenta con una versión gratuita, que ofrece servicios básicos, y con una paga, que facilita servicios *Premium*. De igual forma, esta plataforma está diseñada para ser usada por estudiantes, profesores y personas involucradas en negocios.

V. MARCO METODOLÓGICO

1. *Establecimiento de objetivos*

1.1. *Objetivo general*

Elaborar el *Proyecto CIC-Big Data* para el Centro de Investigación de la Comunicación de la UCAB, el cual está destinado a promover el Periodismo de datos como práctica asociada al Periodismo de investigación.

1.2. *Objetivos específicos*

- Realizar una investigación documental que dé cuenta del surgimiento del Periodismo de datos en el contexto del desarrollo del Big Data.
- Desarrollar contenidos para Big Data a ser empleados en el proyecto.
- Generar una propuesta editorial en línea para promover el Periodismo de datos.

2. *Modalidad de tesis*

De acuerdo con lo establecido en el *Manual del Tesista de Comunicación Social*, el presente trabajo de grado se inscribe dentro de la octava modalidad o pasantía.

Este proyecto es impulsado por el Centro de Investigación de la Comunicación de la Universidad Católica Andrés Bello y tiene como fin principal la creación de un contenido relativo a un mapa digital para la visualización de datos sobre el cine

venezolano. De igual forma, la pasantía incluye la construcción de un producto editorial en línea para la difusión del mapa mencionado.

3. Tipo y diseño de la investigación

El *Proyecto CIC-Big Data* para el Centro de Investigación de la Comunicación de la UCAB se inscribe dentro del tipo de investigación exploratoria.

De acuerdo con Hernández Sampieri, Fernández-Collado y Baptista (2006), esta se realiza cuando “el objetivo es examinar un tema de investigación poco estudiado” (p.101). Para el caso de este trabajo de grado, el Big Data, Open Data y su relación con el Periodismo es aún un tema que requiere mayor exploración y búsqueda de información.

Por otra parte, esta investigación también corresponde con la de tipo documental, la cual es para Palella y Martins (2010), “la que indaga sobre un tema en documentos –escritos u orales- y se concreta exclusivamente en la recopilación de información en diversas fuentes” (p.90). Para la realización del presente proyecto, la investigación con documentos es de gran relevancia y cumple un papel fundamental para su desarrollo.

Con relación al diseño, el de esta tesis-pasantía es de tipo no experimental, ya que la investigación “se realiza sin manipular en forma deliberada ninguna variable” (Palella y Martins, 2010, p.87).

Es importante indicar que este es un proyecto factible. Según el *Manual de Trabajos de Grado de Especialización y Maestría y Tesis Doctorales de la UPEL* (2012) este “consiste en la investigación, elaboración y desarrollo de una propuesta de un modelo operativo viable para solucionar problemas, requerimientos o necesidades de organizaciones o grupos sociales; puede referirse a la formulación de políticas, programas, tecnologías, métodos o procesos” (p.21).

Tomando en consideración lo anterior, el trabajo va dirigido a una universidad y lo que pretende es apoyar la divulgación de una base de datos sobre cine venezolano y su representación gráfica para que tenga mayor provecho desde el punto de vista periodístico, principalmente.

Asimismo, esta tesis se apoya en una serie de pasos que, según el manual de la UPEL, son el diagnóstico, planteamiento y fundamentación teórica de la propuesta; el procedimiento metodológico, actividades necesarias para su ejecución; el análisis y conclusiones, ejecución, y, finalmente, evaluación de la propuesta.

4. Plan de trabajo de la pasantía

Con el fin de dar cumplimiento a cada uno de los tres objetivos específicos de la presente pasantía, se elaboró un cronograma de actividades a realizar entre los meses de octubre de 2014 y agosto de 2015. Asimismo, el *Proyecto CIC-Big Data* fue estructurado en cuatro fases de trabajo:

- **Primera fase:** Se desarrolló en un período de seis meses, entre octubre de 2014 y marzo de 2015. Esta etapa comprendió principalmente la recolección de material bibliográfico relacionado con términos como Big Data, Open Data, Periodismo de datos y de investigación. También, se recopiló la información más importante que sirviera de contexto para la ejecución de las actividades, además se elaboraron resúmenes a emplear en el marco teórico del trabajo de grado.
- **Segunda fase:** Se llevó a cabo entre los meses de abril y junio de 2015. En este bloque se realizó una investigación sobre la teoría de las constelaciones de sentido de Walter Benjamin y se procedió con el manejo de la base de datos de películas venezolanas. A esta se le hizo una revisión general, para luego proceder con la construcción de categorías por géneros y temáticas.

- **Tercera fase:** Se realizó entre los meses de julio y agosto de 2015 y consistió en la creación de una representación gráfica o visualización de las películas venezolanas, previamente clasificadas por géneros y temáticas. Para esta actividad fue necesaria la utilización del sitio web www.mindomo.com, software con el que se pudo crear el mapa virtual.
- **Cuarta fase:** Fue desarrollada a mediados del mes de agosto de 2015 y consistió en la creación de un producto editorial en línea. En este caso, se realizó un sitio web con herramientas de WordPress para publicar el mapa de películas y un breve resumen sobre el proyecto.

Es importante señalar que para llevar a cabo el plan de trabajo del *Proyecto CIC-Big Data* fue acordado con los tutores encargados de asesorar esta investigación un sistema de entregas, realizadas tanto en formato digital como en físico. Por ello, se pautaron reuniones y revisiones quincenales y mensuales, que permitieron el correcto desarrollo de esta pasantía.

VI. ANÁLISIS Y DISCUSIÓN DE LOS RESULTADOS

Un elemento fundamental en el desarrollo de la pasantía fue la data. Partiendo de esto, se trabajó con una base de datos asignada por el CIC-UCAB, la cual contaba originalmente con 406 largometrajes venezolanos. Este material fue elaborado por los estudiantes Mariana Vázquez, Manuel Randolli y David Castillo, quienes cumplen funciones de beca-trabajo en el mencionado centro.

La base de datos fue construida a partir de febrero de 2014 para recoger desde el primer material del que se tiene registro en el país, “Muchachos bañándose en el lago”, del año 1897 y dirigido por Manuel Trujillo Durán, hasta la última película creada para la fecha, “Francisco Massiani: Breve y arbitraria historia de mi vida”, del año 2015 y dirigida por Manuel Guzmán.

Esta base de datos se caracterizó por clasificar la diversa cantidad de material audiovisual venezolano en varios ítems que comprendían: año, título, dirección, guion, ficha técnica, reparto, género, temática, sinopsis, financiamiento, recaudación del año de estreno, recaudación acumulada, espectadores del año de estreno, espectadores acumulados, locaciones, críticas, fuentes, póster e imagen fija. Es importante acotar que la base tomó las categorías y los criterios utilizados por el sitio web IMDb (www.imdb.com).

Asimismo, este proyecto estaba concebido originalmente para que los datos fueran visualizados mediante una interfaz que le permitiera a cualquier usuario navegar en línea. Sin embargo, para efectos del presente trabajo de grado, se acordó simplemente la realización de un demo de la representación gráfica de los largometrajes.

1. Categorización

Para la construcción del mapa de relaciones de películas venezolanas, también conocido como “red neural”, se requirió el establecimiento de una serie de categorías, las cuales estuvieron fundamentadas bajo la idea de la musealización.

En el sitio web www.basterrica.com, Gestión Tecnológica Cultural, se indicó que “musealizar” es un término que hace referencia a “la puesta en valor de los objetos culturales, sitios arqueológicos, monumentos o cualquier otro elemento patrimonial”. De igual manera, se señaló que esto consiste en el diseño del concepto a exponer, para que sea entendible por el receptor o visitante, y para ello necesita incluir un guion lógico de recorrido que fuese atractivo.

Es decir, esta concepción quiere realzar el valor que tienen los materiales culturales, en este caso las películas venezolanas, organizándolos de manera que puedan ser apreciados o comprendidos por la colectividad.

De acuerdo con Blanca Rivero Urdaneta en *Caracas, un museo de arte urbano* (2011), la musealización consiste en “una reflexión que parte de la catalogación y la clasificación y que comprende la categorización del objeto de estudio” (p.88).

Para el caso del *Proyecto CIC-Big Data*, se planteó el desarrollo de una propuesta visual que combinara este principio para que las diversas películas venezolanas estuviesen ordenadas y clasificadas de manera que pudieran resultar atractivas e interesantes para los futuros usuarios.

Por ello, para dar inicio al proceso de categorización, se consideraron solo los ítems de género y temática, ya que estos dos eran mucho más sencillos de clasificar, ayudando así a cumplir con el propósito de conocer qué tipo de cine se ha realizado en el país.

Es importante hacer la salvedad de que al momento de realizar la pasantía, la base de datos se encontraba aún en construcción. Debido a esto, la gran mayoría de las películas no tenían todos los ítems completos, en especial los de género y temática, motivo por el cual se debió realizar un proceso de revisión, compleción y clasificación específicamente para estos dos aspectos.

Asimismo, vale mencionar que algunos de los géneros y temáticas colocados originalmente en la base, como eran: ciencia ficción, testimonial policial, corrupción, thriller, holocausto, crisis bancaria, contrabando y caudillismo; debieron ser modificados al ser demasiado específicos para ser considerados como categorías de temáticas. Sin embargo, sí se tomaron en cuenta para los dos primeros borradores del mapa de relaciones.

1.1. Construcción de las categorías

Como primer paso, se procedió a crear una copia de la base de datos (solo tomando en cuenta el año, título, género y temática) para no modificar o dañar la original. Las 406 películas venezolanas pasaron a ser 386, luego de un proceso de revisión general en el que se descartaron los filmes repetidos y aquellos de los que no se conseguía información suficiente como para poder ser categorizados.

Una vez culminado el proceso mencionado, se empezó a trabajar con la catalogación de los dos ítems seleccionados. Es importante señalar que para esto fue necesaria una búsqueda documental de cada largometraje para determinar exactamente a qué género pertenecía, así como la temática o temáticas que desarrollaba.

1.1.1. Géneros

Luego de una revisión detallada de las sinopsis de cada una de las películas que se incluyeron en la representación gráfica, se pudo determinar siete tipos de géneros

predominantes en la cinematografía nacional. A continuación se presentan en orden decreciente de acuerdo con su frecuencia de aparición:

- **Drama:** Incluye todas aquellas películas con fuerte carga emocional o moral. En este tipo de filmes existe la presencia de situaciones o acciones fuertes, tensas y pasionales. De igual manera, existen giros en la acción o la trama y se desarrolla un conflicto entre protagonistas o entre este y el entorno que le rodea. Dentro de esta categoría se clasificaron doscientos sesenta y siete (267) largometrajes.
- **Documental:** Género no ficcional que representa de manera informativa y didáctica aquellos hechos, escenas y experimentos de la realidad. Tomando en consideración esta idea, cuarenta y ocho (48) filmes de la base de datos fueron categorizados como documentales.
- **Comedia:** Género cinematográfico en el que hay un predominio de aspectos placenteros, festivos o humorísticos, con desenlaces que suelen ser felices. Dentro de este género se incluyeron cuarenta y cuatro (44) de las películas venezolanas.
- **Suspense:** Género que produce expectación, impaciencia, angustia o ansiedad por una acción o escena dramática con final incierto. De las 386 películas con las que se trabajó, solo trece (13) fueron clasificadas dentro de este tipo.
- **Fantasía:** Tipo de cine en el que hay un predominio de la magia, de personajes y criaturas irreales o de mundos fantásticos, los cuales no pertenecen a la realidad que se conoce. Siete (7) largometrajes fueron categorizados como fantásticos.
- **Musical:** Género en el que se desarrollan partes cantadas y bailadas durante la acción. En esta categoría, los filmes son interrumpidos por fragmentos en los que hay canto, baile y coreografías. Solo cuatro (4) largometrajes fueron incluidos.

- **Terror:** Incluye todas aquellas películas cuyas tramas tienen la intención de generar miedo, sobresalto, inquietud, incomodidad, pavor o temor en el espectador. Solamente tres (3) filmes fueron clasificados como terroríficos.

1.1.2. Temáticas

Para la clasificación de las 386 películas venezolanas de acuerdo con sus temáticas, se procedió, al igual que para los géneros, a una revisión y análisis detallado de sus tramas. Es necesario acotar que muchos de estos filmes tratan varias temáticas a la vez, por lo que una misma película podía aparecer en varias clasificaciones.

A continuación se presenta una lista con 29 temáticas que aparecen en el cine nacional, ordenadas de manera decreciente según la cantidad de filmes que estas abarcan:

- **Social:** Incluye todas aquellas películas en las que se presenta una denuncia o crítica de los problemas sociales, hay predominio del uso del vocabulario popular y la trama pretende generar reflexión sobre la sociedad. Partiendo de esta concepción, se clasificaron ciento treinta y cinco (135) películas venezolanas dentro de lo social. Esta temática es la más frecuente en el cine nacional.
- **Delincuencia:** Temática que incluye todos aquellos filmes en los que hay presencia de antisociales y se cometen delitos, asesinatos, robos o se quebranta la ley. Cincuenta y nueve (59) películas venezolanas desarrollaron este aspecto en su trama.
- **Muerte:** En esta temática se incluyen todos los filmes en cuyas tramas el término o cese de la vida es un elemento central o es el desencadenante de nuevas acciones. De las 386 películas con las que se trabajó, cincuenta y siete (57) correspondieron con esta categoría.

- **Romance:** Categoría que se caracteriza por la fuerte presencia de parejas que viven plenamente su amor o que están teniendo una aventura pasajera. Cuarenta y cinco (45) películas venezolanas abarcaron esta temática.
- **Intimista:** Temática característica de los largometrajes que se centran en los asuntos más íntimos de la vida los protagonistas. La cantidad de personajes suele ser muy poca. En el cine venezolano se pudieron encontrar cuarenta y dos (42) películas que desarrollaban este tema.
- **Violencia:** Categoría que incluye todos los filmes en los que se ejerce la fuerza intensa o impetuosa para lograr un determinado objetivo. De igual manera, se presentan dentro de esta categoría las violaciones, golpes, gritos, persecuciones, peleas y asesinatos. Treinta y cinco (35) películas fueron categorizadas como violentas.
- **Policial:** Temática caracterizada por la fuerte presencia de los cuerpos de la policía durante la trama de una película. Vale acotar que la acción gira en torno a este tema. Se categorizaron treinta y cuatro (34) filmes venezolanos como policiales.
- **Familia:** Incluye todas aquellas películas en donde la trama se enfoca en parientes o familiares, o en las que este elemento es de importancia para el desarrollo de la acción. Un total de treinta y un (31) largometrajes desarrollaban esta temática.
- **Pasional:** Esta categoría hace referencia a las relaciones de parejas en las que el sentimiento amoroso no es lo central, sino la pasión. Muchas veces son un medio para lograr un determinado objetivo en la historia; dentro de esta clasificación también se incluyen las infidelidades y las relaciones tormentosas. Veintisiete (27) filmes incorporaron esta temática en sus tramas.

- **Psicología:** Temática en la que los largometrajes están centrados en las conductas humanas o en los aspectos relacionados con la mente: memoria, pensamientos, recuerdos e imaginaciones. Veinticinco (25) filmes venezolanos fueron clasificados en esta categoría.
- **Historia:** Temática de las películas que desarrollan una etapa, época o suceso de relevancia histórica para un determinado país. En el caso venezolano, se clasificaron veinticuatro (24) como históricas.
- **Aventuras:** Temática en la que los sucesos inesperados o con resultados inciertos, que pueden ser riesgosos, se hacen presentes en el desarrollo de la historia. Veintitrés (23) películas se clasificaron en esta categoría.
- **Cultos y religión:** Incluye todas aquellas películas en las que lo religioso o la brujería son centrales o se hacen presentes en la acción. De acuerdo con las sinopsis estudiadas, veintidós (22) películas venezolanas tratan este tema.
- **Música:** Temática en donde la música y los cantos, bailes y coreografías son incluidos en diversos momentos de los largometrajes, interrumpiendo o no la acción. Dentro de esta categoría fueron incluidas veinte (20) películas.
- **Riqueza y ambición:** En esta categoría se incluyeron todas las películas cuyas temáticas están relacionadas con el deseo intenso por el lucro o por la abundancia de bienes. Veinte (20) películas fueron asignadas a esta clasificación.
- **Venganza:** Incluye todos aquellos filmes en los que se puede apreciar una satisfacción personal por parte de algún personaje como resultado del agravio o daño que le produce a otros. Dieciocho (18) películas venezolanas abordaron este tema.
- **Dictadura:** Temática de las películas que hacen especial énfasis en el gobierno dictatorial de un presidente o en las que la presencia de este tipo de gobierno

afecta el desarrollo de la historia. Dieciséis (16) filmes fueron incluidos en esta categoría.

- **Militar:** Temática de aquellos filmes en los que el elemento militar y los guerrilleros tienen un papel importante o central en la acción. Dieciséis (16) de las 386 películas presentaron este tipo de temática.
- **Política:** En esta categoría se encuentran clasificados los filmes en los que el elemento político cobra relevancia o forma parte de la tesis central de la historia. Del total de películas de la base de datos utilizada, dieciséis (16) se categorizaron como políticas.
- **Indígena:** Temática caracterizada por la presencia de personajes, pueblos o tradiciones de naturaleza aborígen durante el desarrollo de la acción. Catorce (14) películas fueron categorizadas como de temática indígena.
- **Poder:** Se incluyen los filmes en los que se evidencia un dominio, imperio o facultad que tiene un personaje para mandar sobre otros o para realizar o ejecutar algo. En el cine venezolano, catorce (14) producciones trataron el poder, el cual es un aspecto de relevancia para la historia.
- **Pobreza:** Temática de las producciones de cine caracterizadas por retratar la escasez de recursos económicos o de elementos necesarios para vivir, que sufre un personaje o una colectividad. En trece (13) películas venezolanas, este tema fue central o relevante para el desarrollo de la historia.
- **Arte:** Incluye las producciones cinematográficas en las que las diversas disciplinas artísticas, como la arquitectura, cine, diseño, fotografía, pintura, escultura, teatro y danza, son centrales o relevantes en la acción. Doce (12) filmes venezolanos tenían esta temática.

- **Biografía:** Dentro de esta categoría se incorporan los largometrajes centrados en recrear la vida de un personaje de importancia, ya sea histórica, social, científica, política, deportiva o artística. Diez (10) producciones se clasificaron en esta temática.
- **Espionaje:** Incluye todos los filmes en los que hay presencia de espías o personas que de manera encubierta tratan de desvelar la identidad de otros personajes, conseguir información importante o, incluso, resolver alguna situación. Siete (7) películas venezolanas desarrollaron esta temática.
- **Parapsicología:** En esta categoría se incluyen todas aquellas producciones cinematográficas en las que se hace presente lo paranormal, las apariciones fantasmales, levitaciones y premoniciones, los cuales aún no son comprobados científicamente. Cinco (5) películas venezolanas han tratado este aspecto en sus historias.
- **Juegos y apuestas:** De la presente categoría forman parte todas aquellas películas que su trama se centra en los juegos de invite y azar, carreras de caballo y apuestas de cualquier tipo. En el cine venezolano se encontraron tres (3) filmes de este tipo.
- **Nacional:** Temática netamente relacionada con la exaltación de lo nacional y de las tradiciones del país. Dos películas venezolanas (2) desarrollan especialmente este tópico en sus historias.
- **Ciencia:** Incluye todos los largometrajes dedicados a la exposición o explicación de fenómenos, investigaciones o logros científicos, ya sea en las áreas de la salud, medicina, naturaleza, ingeniería y astronomía, principalmente. Solo dos (2) filmes son de temática científica.

2. Visualización

La visualización forma parte de las características más resaltantes que definen lo que es el Big Data. Esta cualidad se refiere a la manera de ver o mostrar los datos, lo que permite darles sentido y mayor riqueza, hacerlos entendibles para poder analizarlos, y así obtener un valor agregado de ellos.

Es por ello que para el desarrollo del *Proyecto CIC-Big Data* fue importante plasmar una propuesta visual que representara la base de datos del cine venezolano en un formato atractivo, comprensible y que aportara significado a los distintos filmes y sus respectivas categorías.

Para la visualización de esta data, se tomó como base la teoría de las constelaciones de sentido de Walter Benjamin, la cual pretende establecer “conexiones significativas entre elementos independientes y distantes, y generar una unidad relacional que hable por sí misma” (Círculo de Bellas Artes, 2010). De esta manera, se forma una red neuronal de enlaces.

Lo que se quiso obtener fue una constelación que pusiera de manifiesto las relaciones existentes entre los distintos puntos que la constituyen, en este caso, los vínculos que los filmes venezolanos pueden formar entre sí, para comprender el cine nacional y contribuir con las futuras investigaciones que se hagan al respecto. Todo esto con la ayuda de la visualización.

El modo de presentar los datos en redes neuronales permitió darle significado a la previa categorización de películas que se realizó, así como facilitó la musealización de la información al traducir los datos “invisibles” en imágenes.

El estilo de constelación o red neuronal que forman las conexiones de las películas es visualmente cómodo, condensa una enorme cantidad de información y la

hace manejable, ya que en vez de intentar establecer las similitudes de una larga lista de obras, solo se tienen que observar las relaciones existentes en la constelación.

Además, esta representación permite que los datos queden mejor organizados, lo cual ayuda al usuario a analizar su naturaleza y sentido. Cuatro visualizaciones permitieron que el proyecto avanzara en la aplicación de la teoría de Benjamin. A continuación se explican en detalle:

2.1. Primera propuesta visual

Como resultado de una discusión de trabajo, y luego de seleccionar una muestra de 10 películas al azar, se obtuvo la primera propuesta visual del proyecto. Esta fue elaborada a manera de mapa mental y en formato digital, empleando herramientas de Microsoft Word.

Tres fueron los géneros que se escogieron por considerarse los más representativos: drama, documental y comedia. Además, se trabajó con cuatro temáticas: social, violencia, subjetividad e identidad (ver *Figura 1* en los anexos).

Este modelo pretendía poner de manifiesto las relaciones entre los filmes. Para ello se dispusieron en el espacio, en este caso, las 10 películas seleccionadas de acuerdo con las áreas a las que pertenecían.

Por ejemplo, la película *Araya*, al ser un documental que a la vez trata sobre el tema de identidad, fácilmente podría ser ubicada en el sector documental-identidad. Esto mismo sucede con otras producciones como *Coctel de camarones*, que correspondía a la sección comedia-social.

Esta representación permitió no solo presentar los datos, sino también les aportó un “valor”, para que puedan ser aprovechados. Sin embargo, con este tipo de modelo aún no era visible una red de conexiones propiamente dicha.

2.2. Segunda propuesta visual

La muestra de filmes que se había tomado anteriormente fue ampliada a catorce películas, siete géneros y 10 temáticas. Esta representación gráfica también fue elaborada siguiendo el modelo de un mapa mental en formato digital y usando la herramienta Microsoft Power Point (ver *Figura 2* en los anexos).

Un ejemplo de las relaciones que se obtuvieron fue el caso de la película *Cien Años de Perdón*, dirigida por Alejandro Saderman y estrenada en el año 1998. Este filme perteneció tanto al género de comedia como al de ficción y a la temática de crisis bancaria.

Vale acotar que para el momento de plantear esta propuesta se tomaron en cuenta solo las categorías tal y como aparecían en la base de datos original. Tras el proceso de categorización, la mayoría de las temáticas de las películas cambiaron.

Asimismo, las obras cinematográficas *Compañero de viaje*, considerada dentro de la temática del caudillismo, *Muerte al amanecer*, *Fiebre* y *Simplicio*, clasificadas dentro del cine intimista nacional y *Carmen, la que contaba 16 años*, que trataba sobre el tema del contrabando, pertenecieron al género drama y sirvieron para explicar las conexiones por sectores que se establecieron en este mapa.

Sin embargo, con esta manera de representar la información, la relación entre películas y el concepto de las constelaciones de sentido de Walter Benjamin no quedaba plasmada, ya que ni la estructura de la visualización ni los conceptos que se iban a emplear para construirla (por ejemplo: los géneros y temáticas a los que podía pertenecer una película) estaban definidos, los datos se encontraban desordenados y las interconexiones no se apreciaban con claridad.

2.3. Propuesta final de la visualización

La forma de disponer los datos del Observatorio de Complejidad Económica (proyecto desarrollado por Alexander Simoes, el MIT Media Lab y el Centro para el Desarrollo Internacional de la Universidad de Harvard) sirvió como modelo para la representación gráfica final de las películas venezolanas.

Este observatorio se caracteriza por presentar datos económicos internacionales mediante visualizaciones dinámicas en forma de anillos, pequeños puntos, mapa de árboles, geomapas y redes. Fue esta última la que inspiró la creación de la visualización de los filmes, ya que este diseño enlaza mediante un diagrama de red unidades independientes y distantes entre sí basándose en sus similitudes.

Un ejemplo que ilustra lo anterior es la relación espacio-producto que se realiza con la visualización de redes. Allí existe la categoría piezas-repuestos, la cual a la vez se relaciona con las categorías de lavadoras domésticas, otros productos del hierro y transmisiones.

Todo esto deja ver las correspondencias que estos productos tienen entre sí. Estos puntos forman la constelación de la teoría de Benjamin y tienen que ver con los “elementos resignificados” a los que el filósofo alemán hacía referencia (ver *Figura 3* en los anexos).

Asimismo, este sitio web al no presentar la información como una simple base de datos, sino más bien con diversas representaciones gráficas, facilita la comprensión de la complejidad del sistema económico de una nación y le proporciona un verdadero valor a la data disponible.

La tercera propuesta de visualización del *Proyecto CIC-Big Data* intentó imitar la visualización anteriormente descrita. Este tercer diagrama fue realizado en forma de borrador y en dos láminas de papel bond. Para su elaboración se emplearon las primeras

doscientas películas de la base de datos del CIC-UCAB, las cuales habían sido previamente categorizadas por géneros y temáticas (ver *Figura 4* en los anexos).

En este modelo se colocaron círculos con las 29 temáticas y las películas se conectaban a estos con líneas de colores—de acuerdo con las temáticas a las que pertenecían—, lo que al final produjo un efecto de red neuronal y una mejor comprensión de las relaciones de sentido que posee la cinematografía nacional.

2.4. Mapa de relaciones en digital

La visualización final fue diseñada como un demo con la ayuda del software gratuito Mindomo y para ello se trabajó en total con las 386 películas, así como con los géneros y temáticas definitivos, que resultaron ser siete y 29, respectivamente.

El formato de trabajo que ofreció Mindomo permitió construir un mapa de relaciones en el que los vínculos que compartían las piezas cinematográficas pudieron establecerse con líneas de colores —de acuerdo con sus temáticas—, lo que facilitó la creación de una enorme red neuronal (ver *Figura 5* en los anexos). De igual forma, los títulos de los filmes se colocaron en diversos colores dependiendo de los géneros a los que pertenecían.

Además, se pudo realizar una representación gráfica de los datos con géneros y temáticas que resultó muy atractiva. Con esta propuesta se ayudaría al usuario a navegar más allá de lo tradicional, lo que proporcionaría una mayor comprensión sobre cómo se interrelacionan las películas que forman parte de la producción de cine nacional.

3. *Producto editorial en línea*

Con el fin de cumplir con el tercer objetivo del presente trabajo de grado, se creó un sencillo sitio web que sirvió como herramienta para dar a conocer el mapa de relaciones sobre películas venezolanas (ver *Figura 6* en los anexos). Esta plataforma digital fue creada utilizando la tecnología gratuita de WordPress, por lo que tiene el formato de un blog, siendo su dirección web o URL www.proyectocicbigdata.wordpress.com

El espacio digital creado, que también fue llamado *Proyecto CIC-Big Data*, no solo sirvió de medio para la difusión de la visualización producida, sino, además, como una presentación del resumen del trabajo realizado. El diseño que se le aplicó al blog estuvo inspirado en la musealización, específicamente en la idea de crear un guion lógico que resultara interesante para el usuario.

Por este motivo, el sitio está compuesto por seis pestañas o páginas en las que se encuentra información sobre las distintas fases de la pasantía, lo que serviría para exponerla de manera que pudiese ser comprendida con mayor facilidad. Se creó una especie de museo virtual con el *Proyecto CIC-Big Data*, en el cual el visitante podía realizar un recorrido por los contenidos.

Entre las secciones o páginas creadas se encuentra la de “inicio”, en la que se planteó brevemente en qué consiste el sitio y quiénes lo crearon. La segunda es “el proyecto”, lugar en el que está plasmado el planteamiento del problema de la tesis, así como el objetivo de esta. En tercer lugar está “películas venezolanas”, espacio dedicado a la base de datos de cine nacional y a la presentación de la lista con las 386 películas junto con el año, título, género y temática.

Las últimas tres pestañas son las más importantes del blog, puesto que condensan el trabajo realizado para el proyecto; estas son “categorización”, “la visualización” y

“conclusiones y recomendaciones”. La primera refleja todo el proceso de construcción de las categorías de género y temática y cuáles fueron los criterios empleados para ello.

La siguiente sección muestra la representación gráfica de estos filmes y las propuestas que se generaron hasta llegar al producto final. Por último, existe una página dedicada a las observaciones y conclusiones que se pudieron obtener de la tesis-pasantía.

Un aspecto importante que se tomó en cuenta a la hora de la creación del sitio web fue su apariencia. Seleccionar una plantilla que brindara la sensación de limpieza y organización era fundamental para que los contenidos pudiesen ser localizados con facilidad y comprendidos sin mayores inconvenientes por los usuarios.

Por este motivo, fue empleado el formato “Plane” (disponible gratuitamente en WordPress), el cual le aportó un diseño sencillo al espacio digital. Este último se caracterizó por poseer un menú principal en su parte superior, que incluye un buscador y las páginas mencionadas anteriormente. Asimismo, el blog utilizó una paleta de colores blanco, azul claro y azul oscuro, todo lo cual facilitó la lectura y la navegación.

Es relevante mencionar que el sitio se creó en formato de blog, ya que resultaba económicamente mejor para el CIC-UCAB, quien más adelante podría considerar el adquirir una versión paga. De igual manera, este tipo de sitio es sencillo para el manejo y carga de los contenidos, así como para que los usuarios aporten o sugieran ideas, opiniones y comentarios sobre la filmografía nacional.

Es este último punto, junto al hecho de hacer disponible la información, tanto a los comunicadores sociales como a cualquier persona que así lo requiera, fueron las razones por las que se trabajó bajo el concepto de Open Data o de datos abiertos. Tomando esta idea, el blog serviría como una vía para que la pequeña base con géneros y temáticas, así como su visualización, puedan ser utilizadas libre y gratuitamente.

VII. CONCLUSIONES Y RECOMENDACIONES

El cine venezolano ha producido desde el año 1897 más de 300 obras fílmicas de diversos géneros y temáticas. De entre estas, los dramas sociales son las producciones que más predominan.

Esta conclusión parte del análisis de una representación gráfica realizada sobre películas nacionales. Esta visualización, también conocida como “constelación de sentido” o “red neuronal”, debido a su forma, es el producto principal del presente trabajo de grado, el cual permitió construir un mapa de relaciones útil, interesante y dinámico para ser empleado como herramienta periodística y como medio para dar a conocer esta cinematografía.

En la tesis-pasantía se pudo cumplir con los objetivos específicos previstos, los cuales consistían en una revisión teórica, en la creación de un contenido utilizando Big Data y la generación de un producto editorial en línea. Todo esto con el fin de fomentar la práctica de Periodismo de datos asociada a la de investigación.

Para poder lograr los tres aspectos antes mencionados se requirió el uso de la data, un elemento fundamental que se hizo presente desde el momento de la concepción de este trabajo. Por ello, se contó con una base de datos que alberga 386 películas venezolanas, con información detallada sobre estas, lo que la convierte en un repositorio bien completo sobre el tema.

Con el *Proyecto CIC-Big Data* se evidenció el potencial que representan tanto los datos masivos como los datos abiertos para el Periodismo, ya que se logró obtener información accesible y completa sobre un aspecto no tan abordado como lo es el cine venezolano para ser empleada en futuras investigaciones.

No solo el acceso a grandes cantidades de información sin costo y sin restricciones es fundamental para hacer trabajos periodísticos. También, en la medida en que estos datos tengan una buena representación visual, podrán ser mejor comprendidos para establecer relaciones, profundizar las noticias y crear historias.

Es importante acotar que la visualización es una de las características del Big Data, por lo tanto, este trabajo de grado se enfocó en la construcción de un material basado en esta idea y que permitiera representar gráficamente las diferentes piezas del cine nacional.

Debido a lo anterior, la opción seleccionada para visualizar esa data fue un mapa de relaciones que, basado en la teoría de constelaciones de Walter Benjamin y la musealización, permite reflejar todas las películas venezolanas. De esta forma, un guion lógico, a manera de museo virtual, sirve para relacionar los tipos de filmes y los temas que desarrollan sus historias.

La manera en que los componentes de este material están estructurados ayuda a que la representación gráfica sea atractiva y de sencilla comprensión. Por ello, se disponen pequeños círculos para colocar las temáticas, (a manera de constelaciones), y se emplea una variedad de colores para diferenciar cada género y trama.

Del análisis del demo con la red neuronal se puede concluir que el cine venezolano es rico en dramas, documentales y comedias, siendo los largometrajes de fantasía, suspenso, musical y terror los menos frecuentes. De igual forma, los temas más desarrollados en el país incluyen lo social, la delincuencia, violencia, muerte, romance y lo intimista.

Es importante acotar que la información, al estar en una base de datos conformada por numerosos renglones de categorías, genera dificultad para ser entendida, debido a que no se aprecia cómo esta se relaciona entre sí. En cambio, al estar representada visualmente, los datos se vuelven más sencillos para ser comprendido

porque las conexiones de estos se pueden apreciar a simple vista. De allí la importancia del mapa visual para este proyecto.

Además, la creación del blog permite difundir gratuitamente la base de datos y la representación. De esta manera, el acceso a esta información se hace más sencillo, aplicándose así el Open Data o los datos abiertos.

Aunque la pasantía del *Proyecto CIC-Big Data* culminó por motivos de esta tesis, esto no significa que vaya a ser suspendido. Por el contrario, este trabajo continuará siendo desarrollado por estudiantes, profesores y profesionales del CIC-UCAB. Esto se debe a que el cine venezolano sigue produciendo obras que deberán agregarse a la base de datos y, posteriormente, a la visualización.

Por este motivo, se pretende que el presente trabajo de grado pueda servir como modelo para las nuevas bases de datos que se realicen en el centro. Igualmente, se espera que el producto de esta tesis pueda complementar las investigaciones de los comunicadores sociales venezolanos, así como ayudar a los cineastas a conocer qué tipo de filmes se han estado realizando en el país y qué aspectos o temáticas se podrían abordar en las nuevas producciones.

VIII. FUENTES DE INFORMACIÓN Y BIBLIOGRAFÍA

- **Recursos bibliográficos:**

Brown, B. et al. (2011). *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute.

Corrigan, D. et al. (2013). *Harness the Power of Big Data*. Estados Unidos: McGraw-Hill.

Eaton, C. et al. (2012). *Understanding Big Data*. Estados Unidos: McGraw-Hill.

Espasa. (2005). *Nuevo espasa ilustrado*. España: Espasa

FEDUPEL. (2012). *Manual de Trabajos de Grado de Especialización y Maestría y Tesis Doctorales de la UPEL*. (4ta ed.). Caracas: FEDUPEL.

Ferrer-Sapena, A. y Sánchez-Pérez, E. (2013). Open Data, Big Data: ¿hacia dónde nos dirigimos? *Anuario ThinkEPI*, 7, 150-156.

García C., J. (2013). Infoexplosión. Nuevas estrategias de gestión de la información.

García-Pelayo y Gross, R. (1964). *Pequeño Larousse Ilustrado*. París: Ediciones Larousse. *Telos, Cuadernos de Comunicación e Innovación*, (95), 51-56.

Garriga-Portolà, M. (2011, mayo-junio). ¿Datos abierto? Sí, pero de forma sostenible. *El profesional de la información*, 20 (3), 298-303.

Hernández Sampieri, R.; Fernández-Collado, C. y Baptista, P. (2006). *Metodología de la Investigación*. (4ta ed.). México: McGraw-Hill.

Parella, S. y Martins, F. (2010). *Metodología de la Investigación Cuantitativa* (3era ed.). Venezuela: FEDUPEL.

Paniagua, S. (2013, junio-septiembre). Un mundo de sensores. De los datos al Big Data. *Telos, Cuadernos de Comunicación e Innovación*, (95), 94-96.

Pulido C., E. (2014). *Big Data, ¿solución o problema?* Madrid: Universidad Autónoma de Madrid.

Puyol M., J. (2014). Una aproximación a Big Data. *Revista De Derecho UNED*, (14), 471-505.

Rivero, B. (2012). Caracas: un museo de arte urbano. *Cuadernos de vivienda y urbanismo*. 5 (9), 88-103.

Sherman, C. (2014). *What's the Big Deal About Big Data*, 38 (2), 10-16.

Tapia, A. (s.f.). *Big Data y Periodismo de datos*. Universidad Internacional de La Rioja.

Tascón, M. (2013, junio-septiembre). Dossier, Big Data. *Telos, Cuadernos de Comunicación e Innovación*, (95), 47-50.

- **Recursos electrónicos:**

Abad, M. (2012). *¿Olvidan los periodistas el potencial del Periodismo?* Recuperado el 11 de diciembre de 2014 de <http://www.yorokobu.es/peridosimodatos/>

Ambriz, A. (2013). *Almacenamiento de datos estructurados con Big Data*. Recuperado el 25 de febrero de 2015 de <https://www.ibm.com/developerworks/ssa/library/bd-almacenamiento-datos/#authorN1002D>

Basterrica, Gestión Tecnológica y Cultural. (s.f.). *Musealización*. Recuperado el 7 de junio de 2015 de <http://www.basterrica.com/wordpress/musealizacion/>

Bradshaw, P. (s.f.). Introducción: ¿Qué es el Periodismo de datos? En Gray, J., Bounegru, L., y Chambers, L. (s.f.). *Manual de Periodismo de datos*. Recuperado el 9 de diciembre de 2014 de http://interactivos.lanacion.com.ar/manual-data/introducción_0.html

Chávez, C. (2012). *Los géneros cinematográficos*. Recuperado el 14 de junio de 2015 de http://labutacadecine.blogspot.com/2012/04/los-generos-cinematograficos_07.html

CIC-UCAB. (s.f.). *Centro de Investigación de la Comunicación*. Recuperado el 11 de diciembre de 2014 de <http://cic.ucab.edu.ve/>

Círculo de Bellas Artes. (2010). *Atlas Walter Benjamin. Constelaciones*. Recuperado el 25 de julio de 2015 de http://www.circulobellasartes.com/exposicion/Constelaciones_%285465%29.pdf

Círculo de Bellas Artes/Revista Minerva. (2011). *Walter Benjamin. Constelaciones*. Recuperado el 13 de julio de 2015 de <http://www.revistaminerva.com/articulo.php?id=467>

Círculo de Bellas Artes. (s.f.). *Walter Benjamin. Constelaciones*. Recuperado el 25 de julio de 2015 de http://www.circulobellasartes.com/benjamin_constelaciones/

Cox, M. y Ellsworth, D. (1997). *Application-controlled demand paging for out-of-core visualization*. Recuperado el 20 de octubre de 2014 de <http://www.nas.nasa.gov/assets/pdf/techreports/1997/nas-97-010.pdf>

Crucianelli, S. (2013). ¿Qué es el periodismo de datos? *Cuadernos para periodistas* (26), 106-124. Recuperado el 9 de diciembre de 2014 de <http://www.apmadrid.es/images/stories/106-124%281%29.pdf>

Delbueno, H. (2010). Walter Benjamin, redimiendo el materialismo histórico para una praxis revolucionaria. *Fundamentos en Humanidades*, 2 (22), 35-46. Recuperado el 19 de junio de 2015 de <http://www.redalyc.org/articulo.oa?id=18419812002>

Díaz-Struck, E. (s.f.). Datos que potencian la investigación: El caso de Venezuela. *Manual de Periodismo de datos iberoamericano*. Recuperado el 10 de diciembre de 2014 de <http://manual.periodismodedatos.org/emilia-diaz-struck2.php>

Diebold, F. (2000). *Big Data Dynamic Factor Models for Macroeconomic Measurement and Forecasting*. Recuperado el 20 de octubre de 2014 de <http://www.ssc.upenn.edu/~fdiebold/papers/paper40/temp-wc.PDF>

IBM Big Data and Analytics Hub. (2013). *The four V's of Big Data*. Recuperado el 23 de octubre de 2014 de <http://www.ibmbigdatahub.com/infographic/four-vs-big-data>

IBM. (s.f.). *¿Qué es Big Data?* Recuperado el 10 de noviembre de 2014 de <http://www-01.ibm.com/software/ve/data/bigdata/index.html>

IBM100. (s.f.). *Relational Database*. Recuperado el 11 de agosto de 2015 de <http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/reldb/>

Jameson, F. (2011). Lecturas de Benjamin. *Revista Minerva*. (17ma ed.). Recuperado el 25 de julio de 2015 de <http://www.revistaminerva.com/articulo.php?id=474>

John, O. (2010). *Emancipación del valor de cambio respecto al valor de uso. Reflexiones para una teoría de la Mercancía en la obra de Los Pasajes*. Recuperado el 25 de julio de 2015 de http://www.constelaciones-rtc.net/02/CRTC_02_2010.pdf

Kayser-Bril, N. (s.f.). Introducción: ¿Por qué es importante el Periodismo de datos? Una respuesta a las relaciones públicas basadas en datos. En Gray, J., Bounegru, L., y Chambers, L. (s.f.). *Manual de Periodismo de datos*. Recuperado el 9 de diciembre de 2014 de http://interactivos.lanacion.com.ar/manual-data/introducción_2.html

La Nación. (2012). *La nueva revolución de la información: El periodismo de datos*. Recuperado el 10 de diciembre de 2014 de <http://www.lanacion.com.ar/1493532-la-nueva-revolucion-de-la-informacion-el-periodismo-de-datos>

Lavert Proyectos Culturales. (s.f.). *Musealización*. Recuperado el 7 de junio de 2015 de <http://www.lavert.net/Expertos-en/Musealizacion>

Lorente, J. (2012). *Arte público en Internet: Precedentes y novedades del catálogo online en la web municipal de Zaragoza*. Recuperado el 7 de junio de 2015 de http://www.academia.edu/2014528/Arte_p%C3%BAblico_en_internet_precedentes_y_novedades_del_cat%C3%A1logo_on_line_en_la_web_municipal_de_Zaragoza

Lorenz, M. (s.f.). Introducción: ¿Por qué debieran usar datos los periodistas? En Gray, J., Bounegru, L., y Chambers, L. (s.f.). *Manual de Periodismo de datos*. Recuperado el 9 de diciembre de 2014 de http://interactivos.lanacion.com.ar/manual-data/introducci%C3%B3n_1.html

Manchado, M. (2014). *La "Escuela de Frankfurt", una constelación en movimiento*. Recuperado el 14 de julio de 2015 de http://www.scielo.org.ar/scielo.php?script=sci_arttext&pid=S166856282014000100001&lng=es&nrm=iso

Mariño, A. (2012). *La relación del Periodismo de datos con Big Data y Open Data. Trabajo fin de Master en Investigación en Documentación*. Recuperado el 11 de diciembre de 2014 de <http://es.scribd.com/doc/111830154/La-relacion-del-Periodismo-de-Datos-con-Big-Data-y-Open-Data>

Meyer, P. (s.f.). Introducción: ¿Por qué es importante el Periodismo de datos? Filtrar el flujo de datos. En Gray, J., Bounegru, L., y Chambers, L. (s.f.). *Manual de Periodismo de datos*. Recuperado el 9 de diciembre de 2014 de http://interactivos.lanacion.com.ar/manual-data/introducción_2.html

Mindomo. (s.f.). *Mindomo About*. Recuperado el 7 de agosto de 2015 de <https://www.mindomo.com/es/about/mindomo-about.htm>

Morán, I. (2014). *En Venezuela no tenemos una ley de acceso a la información pública*. Recuperado el 9 de diciembre de 2014 de http://www.agenciadenoticias.luz.edu.ve/index.php?option=com_content&task=view&id=4932&Itemid=176

Mundo Contact. (2014). *México y Brasil lideran la inversión en Big Data en AL*. Recuperado el 10 de noviembre de 2014 de <http://mundocontact.com/mexico-y-brasil-lideran-la-inversion-en-big-data-en-al/#sthash.6LLbKYE0.dpuf>

Observatorio de Complejidad Económica (s.f.). Recuperado el 19 de marzo de 2015 de <http://atlas.media.mit.edu/>

Open Data. Portal de dades obertes (s.f.). *Open Data de los ayuntamientos de Cataluña*. Recuperado el 10 de agosto de 2015 de <http://opendata.cloudbcn.cat/MULTI/es>

Osores, M. (2015). *Crece Big Data en AL en 2015 impulsada por software y servicios*. Recuperado el 25 de abril de 2015 de <http://searchdatacenter.techtarget.com/es/cronica/Crece-big-data-en-AL-en-2015-impulsada-por-software-y-servicios>

Paniagua, S. (2012). *Big Data en sanidad para predecir, prevenir y personalizar*. Recuperado el 16 de febrero de 2015 de <http://www.sorayapaniagua.com/2012/11/12/big-data-en-sanidad-para-predecir-prevenir-y-personalizar/>

Pérez Arbesú, L. (2013). *Big Data en América Latina avanza a pasos pequeños*. Recuperado el 10 de noviembre de 2014 de <http://searchdatacenter.techtarget.com/es/cronica/Big-Data-en-America-Latina-avanza-a-pasos-pequenos>

Pérez Arbesú, L. (2015). *Crece lentamente la relevancia del almacenamiento en América Latina*. Recuperado el 10 de noviembre de 2014 de <http://searchdatacenter.techtarget.com/es/cronica/Crece-lentamente-la-relevancia-del-almacenamiento-en-America-Latina>

Pilhofer, A. (s.f.). Introducción: ¿Por qué es importante el Periodismo de datos? Nuevos enfoques para narrar historias. En Gray, J., Bounegru, L., y Chambers, L. (s.f.). *Manual de Periodismo de datos*. Recuperado el 9 de diciembre de 2014 de http://interactivos.lanacion.com.ar/manual-data/introducción_2.html

Press, G. (2013). Recuperado el 18 de marzo de 2015 de <http://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/>

Producción Multimedial Digital UNaM. (28 de agosto de 2012). *Características básicas del Periodismo de datos/Sandra Crucianelli* [Archivo de video]. Recuperado el 9 de diciembre de <https://www.youtube.com/watch?v=HtnrOVuefml>

RAE, Diccionario de la Real Academia Española. (s.f.). Recuperado el 20 de julio de 2015 de <http://www.rae.es/>

Rendueles, C. y Useros, A. (2010). *Atlas Walter Benjamin. Instrucciones de uso*. Recuperado el 25 de agosto de 2015 de <http://es.scribd.com/doc/61023881/Atlas-Constelaciones-Walter-Benjamin#scribd>

S.A. (2009). *Cine intimista*. Recuperado el 25 de junio de 2015 de http://www.lisadamus.com/cine_intimista.html

S.A. (2005). *Datos simples semiestructurados*. Recuperado el 21 de julio de 2015 de http://www.lisadamus.com/cine_intimista.html

Saber UCAB, Repositorio Institucional de la Universidad Católica Andrés Bello. (s.f.). Recuperado el 21 de junio de 2015 de <http://saber.ucab.edu.ve/>

Synergic partners. (s.f.). *Visualizar para entender*. Recuperado el 15 de abril de 2015 de <http://www.synergicpartners.com/que-hacemos/disciplines/visualizacion-datos/>

Universidad Católica Andrés Bello. (s.f.). *Centro de Investigación de la Comunicación*. Recuperado el 11 de noviembre de 2014 de http://w2.ucab.edu.ve/inicio_cic.html

Vermanen, J. (s.f.). Introducción: ¿Por qué es importante el Periodismo de datos? Actualizar sus capacidades. En Gray, J., Bounegru, L., y Chambers, L. (s.f.). *Manual de Periodismo de datos*. Recuperado el 9 de diciembre de 2014 de http://interactivos.lanacion.com.ar/manual-data/introducción_2.html

Vidal Gil, J. (2014). *Big data: Gestión de datos no estructurados*. Recuperado el 15 de febrero de 2015 de <http://www.dataprix.com/blog-it/big-data/big-data-gestion-datos-no-estructurados>

Winshuttle. (s.f.). *Big Data y la historia del almacenamiento de la información*. Recuperado el 12 de agosto de 2015 de <http://www.winshuttle.es/big-data-historia-cronologica/>

IX. ANEXOS

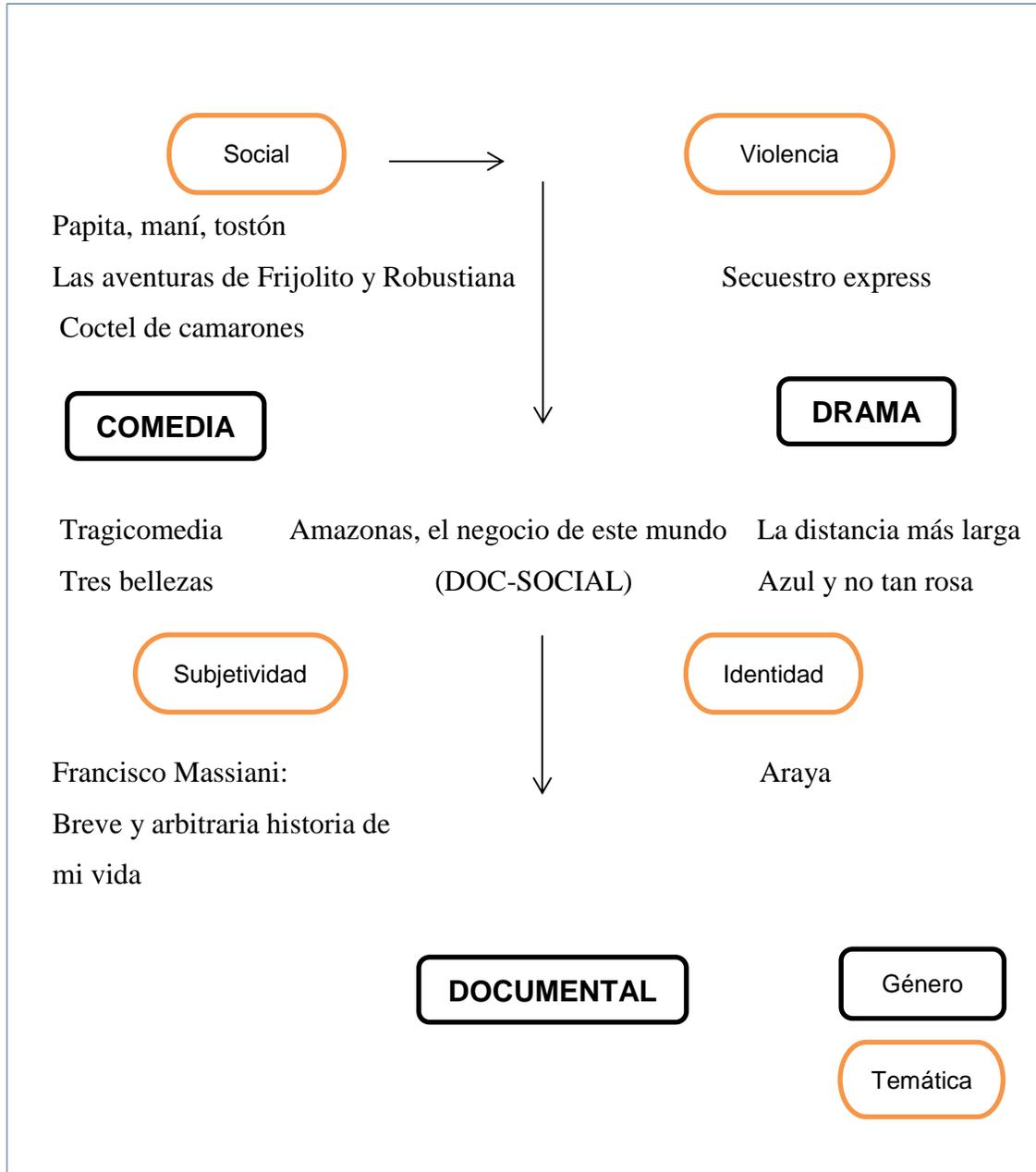


Figura 1. Primera propuesta de representación de películas venezolanas por género-temática

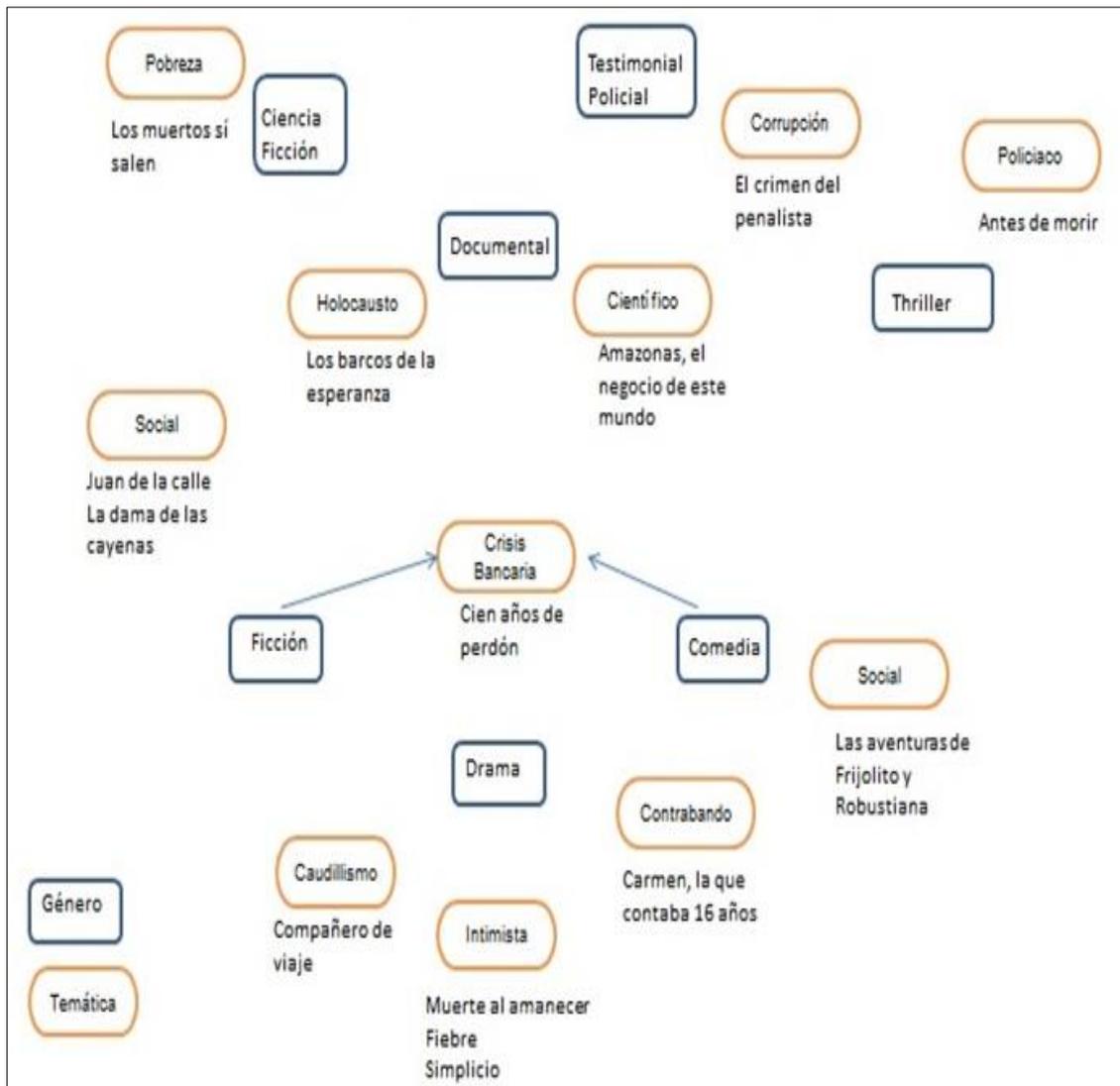


Figura 2. Segunda propuesta de representación de películas venezolanas por género-temática



Figura 3. Relación espacio-producto de las exportaciones de Venezuela (2012) - Observatorio de Complejidad Económica

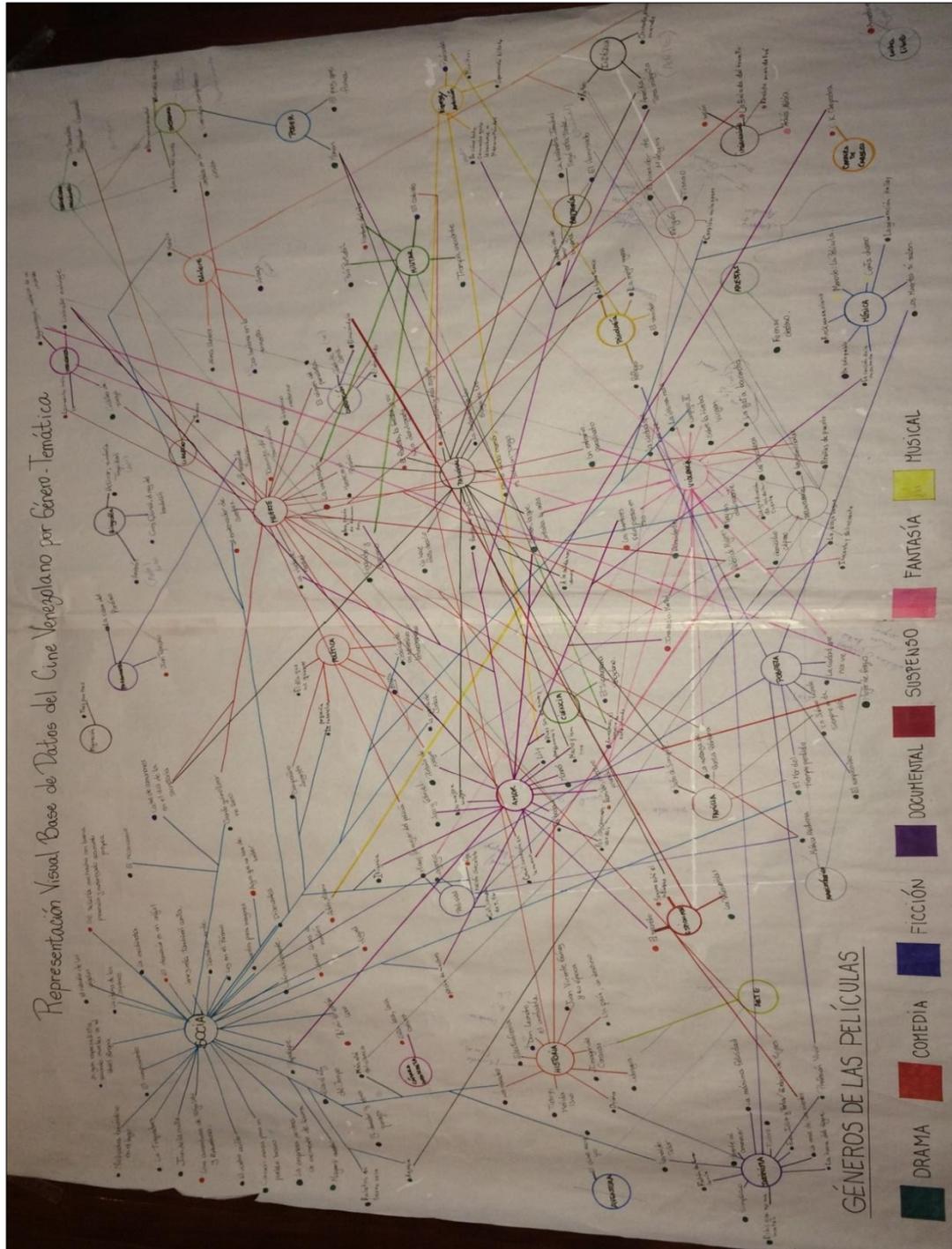


Figura 4. Borrador final en papel del mapa de relaciones con 200 filmes por género-temática

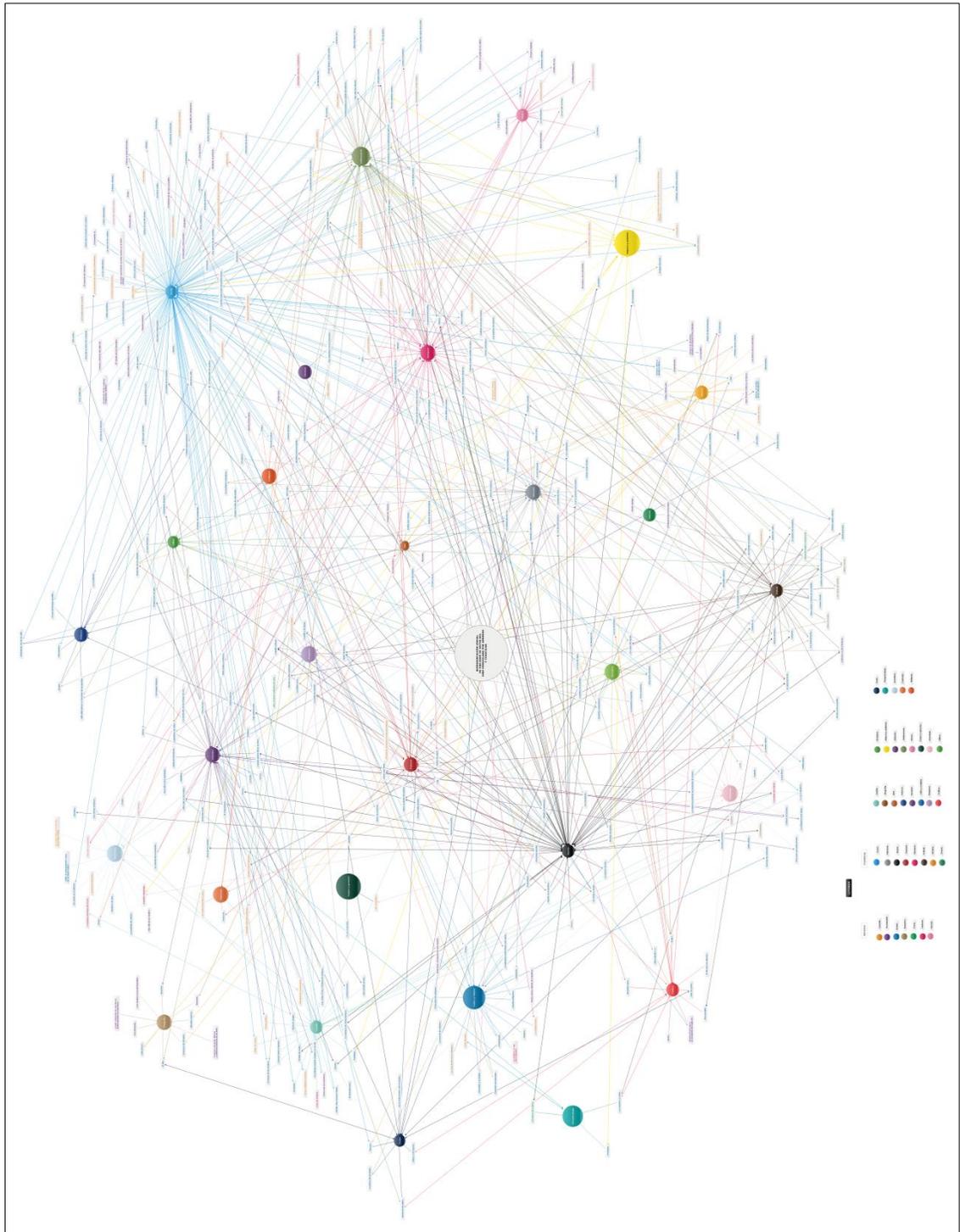


Figura 5. Mapa de relaciones de películas en formato digital



Figura 6. Home o portal del sitio web Proyecto CIC-Big Data