



# CURSO ANÁLISIS DE BIG DATA



Rosario Rogel-Salazar  
Universidad Autónoma del Estado de México  
0000-0002-6018-0635



Verónica Benítez-Pérez  
Universidad Autónoma del Estado de México  
0000-0003-4848-2229





# Datos abiertos

Si este recurso te es de utilidad puedes usarlo libremente, con la única restricción de citarlo como se indica a continuación:

**Rogel-Salazar, Rosario y Benítez-Pérez Verónica (2019), “Curso: Análisis de BigData”, impartido en DIDEPA, Universidad Autónoma del Estado de México, 1 al 5 de julio del 2019.**

DOI: [10.6084/m9.figshare.8796305](https://doi.org/10.6084/m9.figshare.8796305)





# TEMA 1.

## Web 1,2,3 y 4



# Web 1, 2, 3 y 4



## WEB 1.0

**1990**

Solo permitía consumir contenido, pero sin interactuar

Estática, centralizada, secuencial, de solo lectura y no es interactiva.

Correo, navegadores, motores de búsqueda.



## WEB 2.0

**2004**

**Web social:** basada en comunidades de usuarios: foros, blogs, comentarios y redes sociales.

Permite compartir información y colaborar: wikis



## WEB 3.0

**2010**

**Web semántica:**, agrega metadatos semánticos y ontológicos

Recurre a data-web, permite big-data.



## WEB 4.0

**2016**

**Deep Learning / Machine Learning**

Siri, Google Now, Cortana,

Recurre a la AI

Tránsito de la web que da información a la que da soluciones.

Cómputo Cognitivo





# Web 2

## TEMA 1



**FORTUNE 500** 1955-2005 ▼  
 A database of 50 years of FORTUNE's list of America's largest corporations

View by year: 1955 ▼ View by company: A ▼

Full List Companies Profits Assets Current FORTUNE 500

1955 Full list Current View: 1-100 ▼

Rank	Company	Revenues (\$ millions)	Profits (\$ millions)
1	<a href="#">General Motors</a>	9,823.5	806.0
2	<a href="#">Exxon Mobil</a>	5,661.4	584.8
3	<a href="#">U.S. Steel</a>	3,250.4	195.4
4	<a href="#">General Electric</a>	2,959.1	212.6
5	<a href="#">Esmark</a>	2,510.8	19.1
6	<a href="#">Chrysler</a>	2,071.6	18.5
7	<a href="#">Armour</a>	2,056.1	1.6
8	<a href="#">Gulf Oil</a>	1,705.3	182.8
9	<a href="#">Mobil</a>	1,703.6	183.8
10	<a href="#">DuPont</a>		44.4

Full List Companies Profits Assets Current FORTUNE 500

1995 Full list Current View: 1-100 ▼

Rank	Company	Revenues (\$ millions)	Profits (\$ millions)
1	<a href="#">General Motors</a>	154,951.2	4,900.6
2	<a href="#">Ford Motor</a>	128,439.0	5,308.0
3	<a href="#">Exxon Mobil</a>	101,459.0	5,100.0
4	<a href="#">Wal-Mart Stores</a>	83,412.4	2,681.0
5	<a href="#">AT&amp;T</a>	75,094.0	4,676.0
6	<a href="#">General Electric</a>	64,687.0	4,726.0
7	<a href="#">Intl. Business Machines</a>	64,052.0	3,021.0
8	<a href="#">Mobil</a>	59,621.0	1,079.0
9	<a href="#">Sears Roebuck</a>	54,559.0	1,454.0
10	<a href="#">Altria Group</a>	53,776.0	4,725.0

### The 10 Most Valuable Companies on the Fortune 500

While Walmart is king on the Fortune 500, which ranks by revenue, investors see more promise in tech firms—bestowing Silicon Valley with some of the highest market values out there.

Value Rank ▼	Company	Market Value	Sector	Revenues (Bil)	F500 Rank
1	Apple	\$921 bil.	Technology	\$229 bil.	4
2	Amazon.com	\$765 bil.	Retailing	\$178 bil.	8
3	Alphabet	\$750 bil.	Technology	\$111 bil.	22
4	Microsoft	\$746 bil.	Technology	\$90 bil.	30
5	Facebook	\$531 bil.	Technology	\$41 bil.	76
6	Berkshire Hathaway	\$492 bil.	Financials	\$242 bil.	3
7	JPMorgan Chase & Co.	\$388 bil.	Financials	\$114 bil.	20
8	Exxon Mobil	\$349 bil.	Energy	\$244 bil.	2
9	Johnson & Johnson	\$332 bil.	Health Care	\$77 bil.	37
10	Bank of America Corp.	\$315 bil.	Financials	\$100 bil.	24



As of 5/21/2018

SOURCE: Fortune 500

FORTUNE



- Web 3

# Masiva difusión de las tecnologías digitales





TEMA 1

¡ En un segundo...!



1,700 aplicaciones



44,000 búsquedas



1,800 publicaciones



300,000 mensajes WhatsApp



1,700 llamadas

NETFLIX

655 horas de video



TEMA 1

# Economías digitales



# Minimización

# FORMACIÓN





TEMA 1

# ECONOMÍA MUNDIAL: GLOBAL Y DIGITAL



<https://didepa.uaemex.mx>

CURSO "ANÁLISIS DE BIG DATA"

1 al 5 de julio 2019

Rosario Rogel-Salazar / Verónica Benítez-Pérez





TEMA 1

- En 2015:

4,700 millones de personas con suscripción a telefonía móvil

3,174 millones de habitantes (43.4% de la población) usan Internet

4,200 millones de suscripciones a banda ancha fija y móvil

72.500 petabytes al mes tráfico de IP

179,600 millones de descargas de aplicaciones, es decir, 25 por habitante





## TRÁFICO DE DATOS DIGITALES

### Difusión de las tecnologías digitales en el mundo, 2015

4 700 millones de suscriptores de telefonía móvil

3 174 millones de usuarios de Internet

4 200 millones de suscriptores de banda ancha

72 500 petabytes mensuales de tráfico IP

179 600 millones de aplicaciones descargadas

Fuente: Unión Internacional de Telecomunicaciones (UIT), *ICT Indicators Database*, 2016; GSMA, *The Mobile Economy 2015*, 2015; y Statista, *The Statistical Portal*.



### Uso de aplicaciones y actividades en Internet a nivel mundial, 6 de julio de 2015



Fuente: Comisión Económica para América Latina y el Caribe (CEPAL), sobre la base de Internet Live Stats [en línea] <http://www.internetlivestats.com/>.



# Patrón de consumo digital

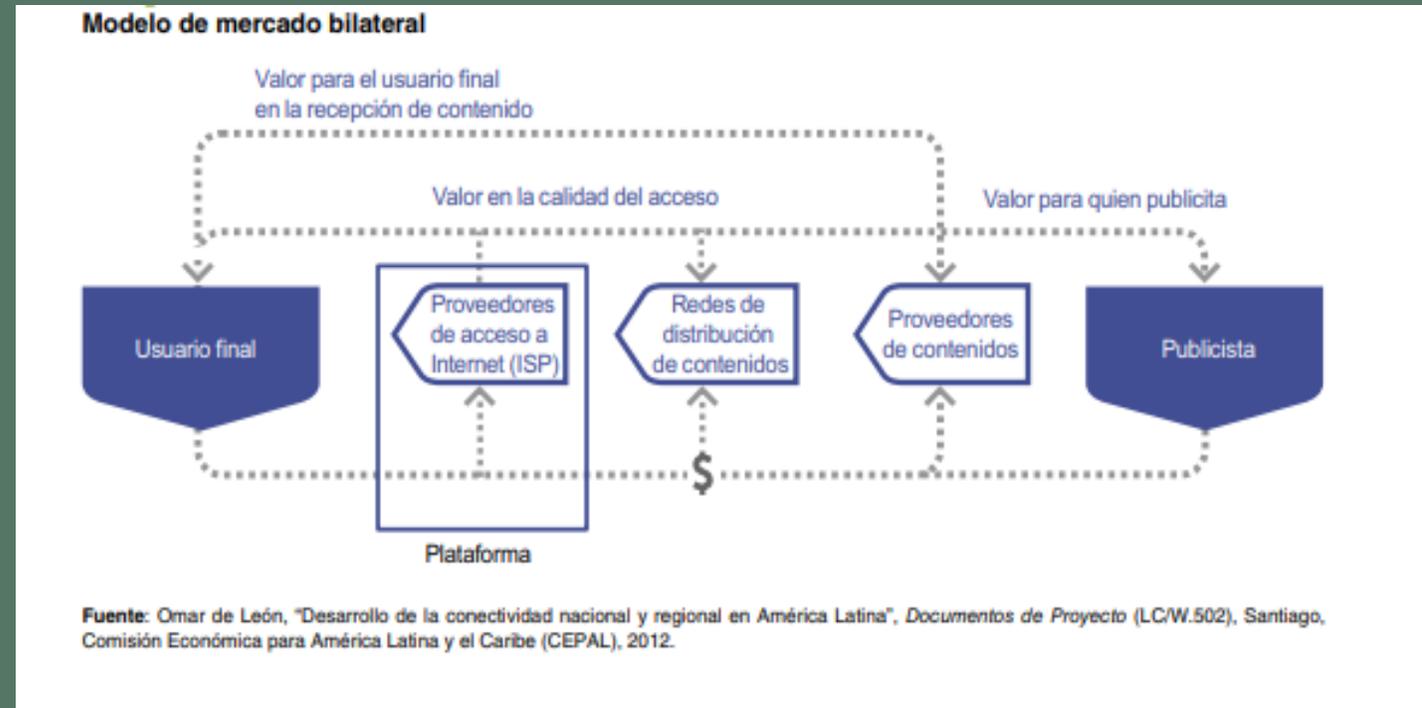
- **ISP:** permiten acceso a la red, para descargar y utilizar contenidos, aplicaciones y servicios.
- **Mercado unilateral:**
  - Proveedores Usuarios
  - Intermediarios de proveedores





TEMA 1

# Neutralidad de la red: Blockchain





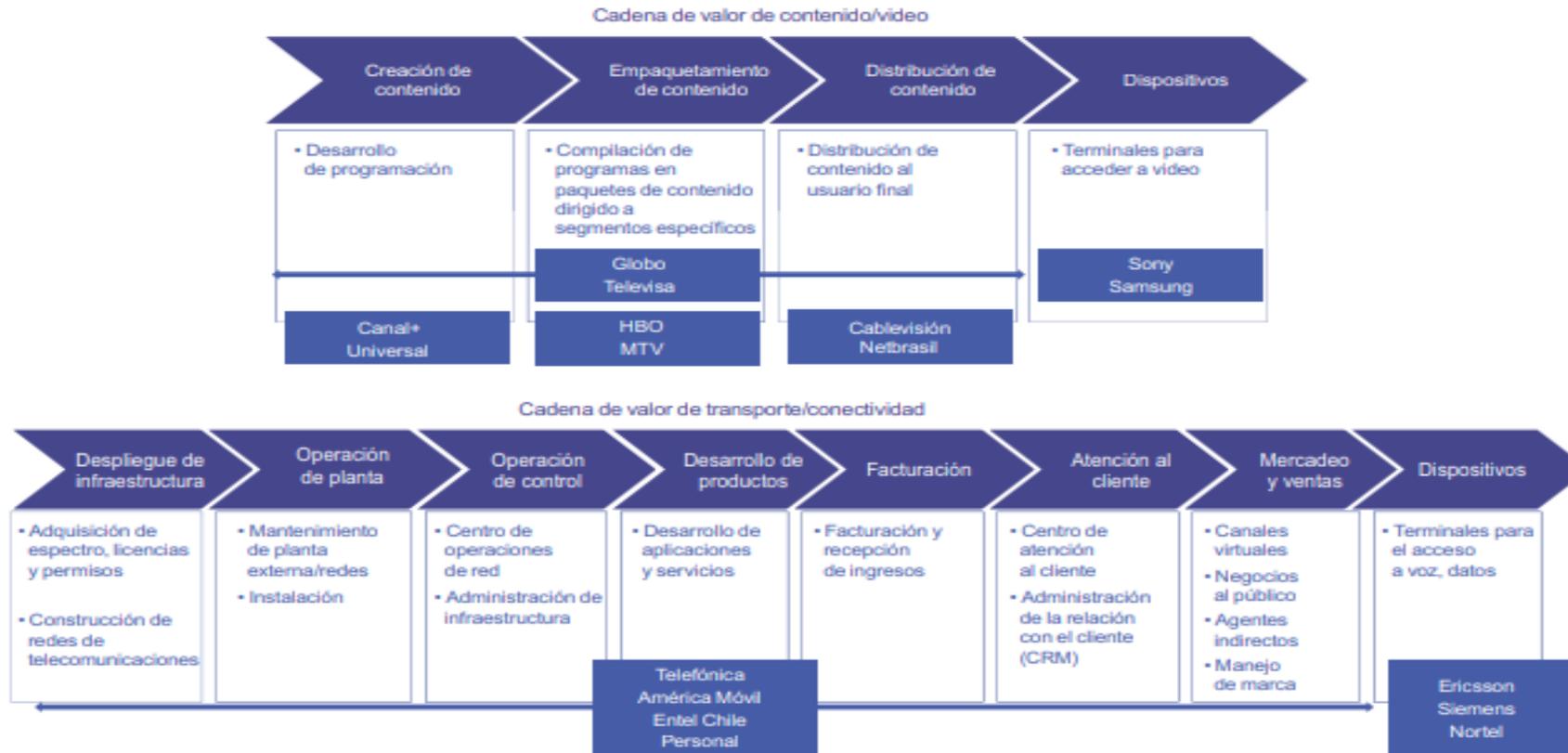
# Huella digital

Cuando utilizas un servicio y no tienes que pagar,  
es porque el **producto eres tú.....**



# Web 4

## Cadenas de valor originales



Fuente: Telecom Advisory Services (TAS).

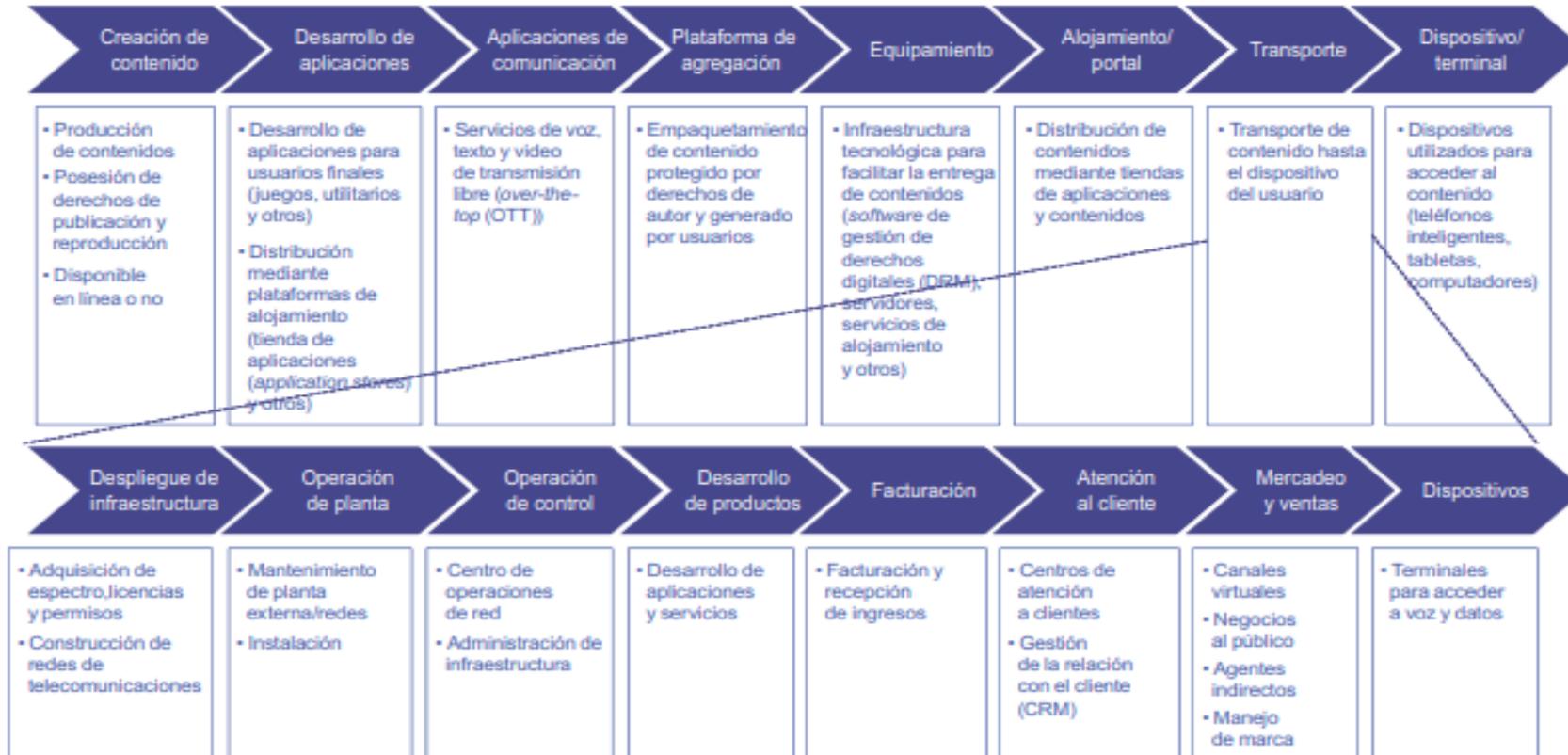




# Web 4

TEMA 1

### Cadena de valor de contenidos y servicios digitales



Fuente: Telecom Advisory Services (TAS).





# Web 4

TEMA 1

## Cadena de valor máquina a máquina (M2M)



Fuente: Cartesian, *Accelerating Monetization of M2M/Connected Devices*, 2013.





TEMA 2

# Tema 2. ¿Qué es el Big Data?

Dato del latín *dare*: dar

Dato: siglo XVII

Big data año 2000

Valor económico y político





# Atributos técnicos del Big Data

Volumen : cantidad de datos

1,114 terabytes

16 millones de fotografías  
de Facebook= 1 terabyte





# Atributos técnicos del Big Data

Velocidad:

Flujo de la generación de datos

No. De individuos que pasan  
controles en aeropuerto, registros  
de vigilancia



TEMA 2

COMPONENTES  
DEL BIG DATA





TEMA 2

El nombre completo de “Big Data” es  
“Análisis de Big Data”:

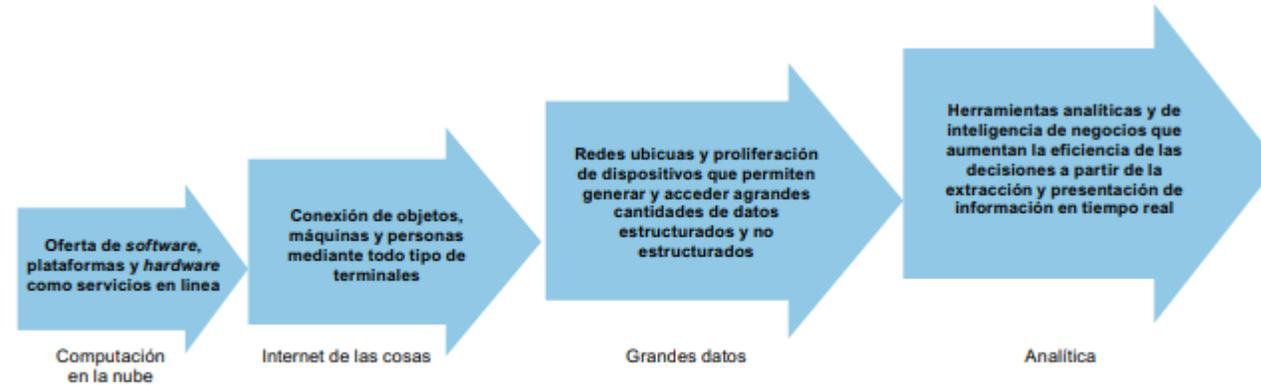
- Flujo de información
- Almacenamiento de la información
- Información computacional





# CARACTERÍSTICAS DEL BIG DATA

## De la computación en la nube a la analítica de los grandes datos



Fuente: Comisión Económica para América Latina y el Caribe (CEPAL), sobre la base de Frost & Sullivan's Global Digital Media Research, "Internet of Things, Big Data and Analytics 101" [en línea] <http://es.slideshare.net/mukulkrishna/internet-of-things-and-big-data-101>.



TEMA 2

# Flujo de información

2 páginas de periódicos por persona por día en 1986 (0,3 exabytes óptimamente comprimidos en todo el mundo, 20%).

Digitalizado: a seis periódicos completos dos décadas más tarde en 2007 (65 exabytes en todo el mundo, 99.9% digitalizado)

(Hilbert y López, 2011;  
Hilbert, 2011a).

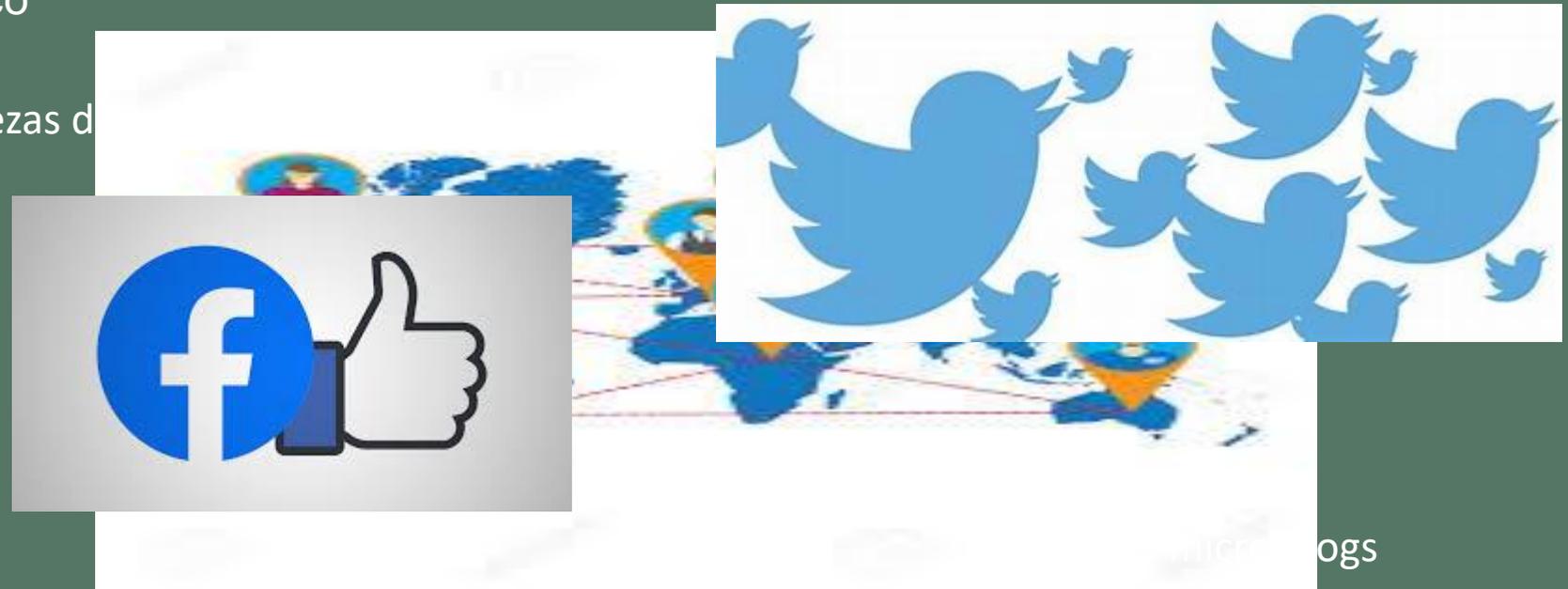




TEMA 2

## LÍDERES EN TÉRMINOS DE USUARIOS EN FACEBOOK EN 2015: India, Brasil, Indonesia y México

700,000 piezas d



(Statista, 2014)



TEMA 2

# TRAZABILIDAD DE LA INFORMACIÓN

- Huellas humanas embebidas en satélites scanners
- Redes sociales digitales
- Energía social en Twitter







TEMA 2

# Información computacional

Educación



Transacciones bancarias



Salud

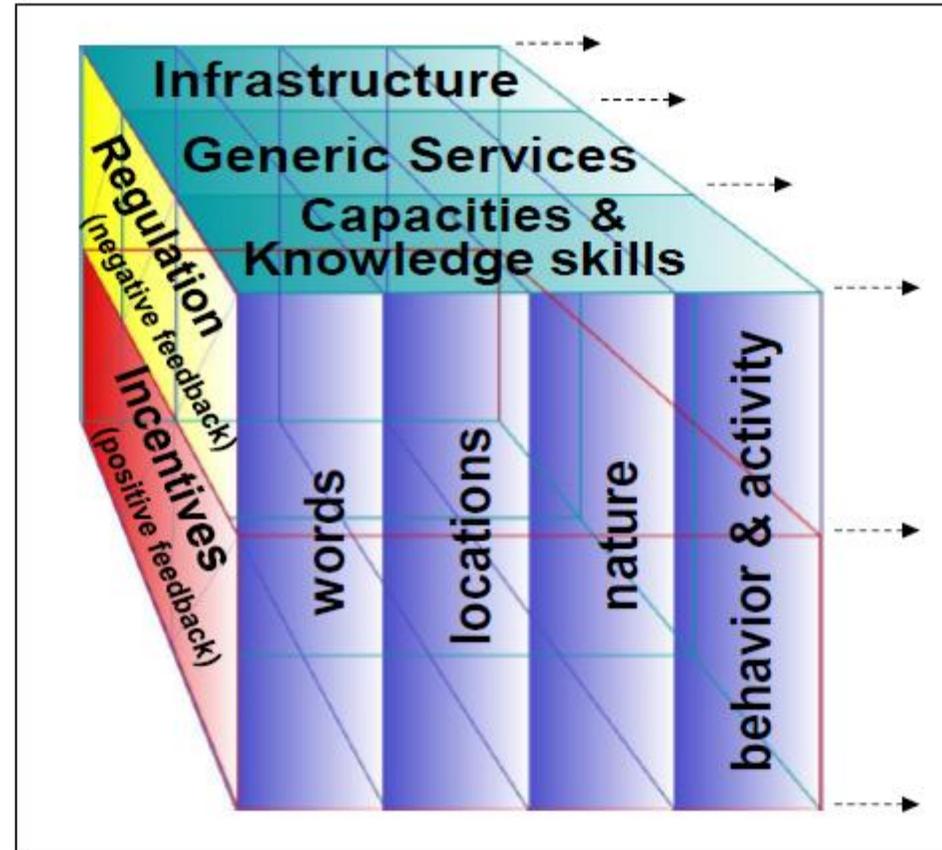




# Desarrollos TIC

## TEMA 2

The three-dimensional "ICT-for development-cube" framework applied to Big Data.





# BIG DATA

## ¿Qué es?

Práctica de combinar inmensos volúmenes de información, proveniente de diferentes fuentes, analizarlos y con frecuencia usar algoritmos que pueden seguir aprendiendo por su cuenta para ayudar a la toma de decisiones.



### 3 rasgos distintivos de BIG DATA



VOLUMEN



VELOCIDAD



VARIEDAD

### BENEFICIOS DEL USO DEL BIG DATA

1

OFRECER UNA VISIÓN MÁS PRECISA DE LAS FLUCTUACIONES Y RENDIMIENTOS DE TODO TIPO DE RECURSOS.

2

ACCELERAR LA INNOVACIÓN Y LA PRESTACIÓN DE SERVICIOS CADA VEZ MÁS INNOVADORES Y MÁS EFICIENTES.

### RIESGOS DEL USO DEL BIG DATA

1

CAER EN CONCLUSIONES ERRÓNEAS QUE NADIE REvisa.

2

TOMAR DECISIONES AUTOMATIZADAS SIN UN SESGO HUMANO



Asociación de Internet.mx

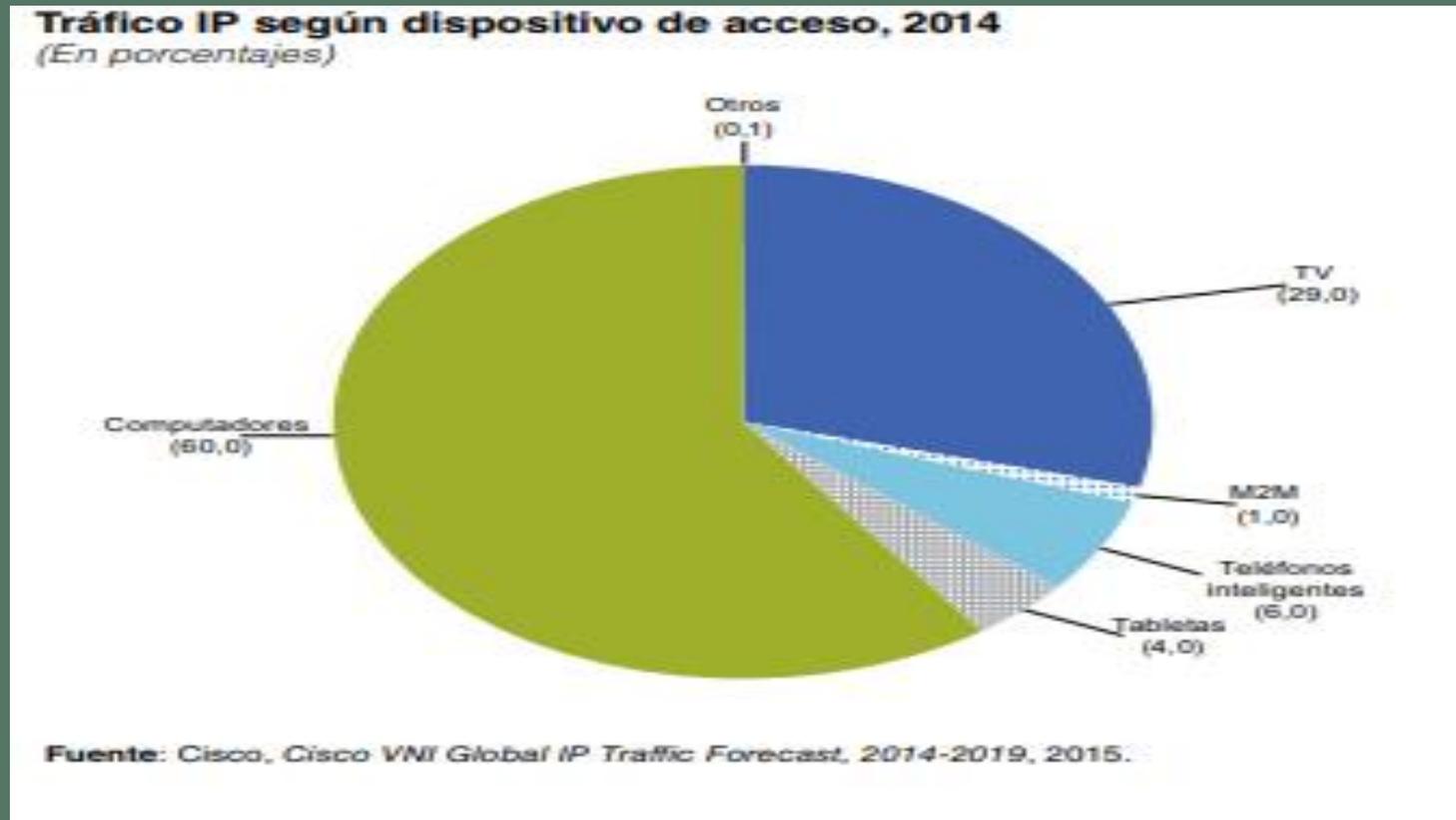
- ✓ Combinar
- ✓ Multivariados
- ✓ Algoritmos
- ✓ Rendimientos
- ✓ Innovación

- ❖ Errores
- ❖ Sesgos



TEMA 2

# DISPOSITIVOS Y SU ACCESO





# CARACTERÍSTICAS CUALITATIVAS

1. Big Data se produce siempre
2. Big Data reemplaza el muestreo aleatorio
3. Big Data siempre está accesible en tiempo real
4. Big Data combina diferentes fuentes
5. Big Data es siempre análisis



TEMA 3

# Tema 3. Uso de plataformas de apoyo a la investigación

- Asimetrías en la generación y transmisión del conocimiento
- El entorno digital: herramienta para construir conocimiento
- Brechas digitales y cognitivas
- Explosión de información





# Comunicación científica. etapas

**1665:** fundación de las primeras revistas científicas (Philosophical Transactions y Journal de Sçavans) y consolidación del sistema científico moderno.

**1945 en adelante:** sedimentación de la “ciencia de corriente principal” a través de bases de datos (Institute for Scientific Information / Web of Science):

Generación de políticas científicas con sentido utilitarista.

Evaluación de la calidad de la ciencia a través de parámetros cuantitativos (factor de impacto, patentes, formación de recursos humanos).

Aumento indiscriminado de precios de las publicaciones.

**2010 en adelante:** en el marco del entorno digital:

Movimiento Acceso Abierto (Open Access)

Colaboración científica a través de la apertura (openness) a la información (Open Science).





TEMA 3

## Redes sociales horizontales y verticales



**Generalistas:** Facebook, Twitter, Google Plus, Tagged, Sina Weibo, Qzone

**De contacto y citas:** Meetic, Badoo, Match, Tinder, Ginder

**Agregador de noticias:** Reddit, Delicious, Menéame

**Turismo y viajes:** Tripadvisor, TopRural, Minube

**Salud y ejercicio:** HealthTrap, Endomondo, Runkeeper

**Foto y video:** Youtube, Instagram, Pinterest, Flickr, Vimeo, Tango, Imgur

**Mensajería:** Whatsapp, FB Messenger, Telegram, Hangout

**Música:** Soundcloud, Spotify, LastFM, MySpace

**Negocios:** LinkedIn

**Academia:** ResearchGate, Academia.Edu, Google Scholar

Rogel-Salazar, Rosario (2019), "Estrategias de visibilidad de la producción académica de las politólogas en medios digitales" conferencia presentada en el Instituto de Investigaciones Jurídicas, Universidad Nacional Autónoma de México, organizada por la Red de Politólogas, viernes 14 de junio del 2019. DOI: [10.6084/m9.figshare.8279720](https://doi.org/10.6084/m9.figshare.8279720)





## Métricas para la Ciencia 2.0

### The Colors of the Donut

- Policy documents
- News
- Blogs
- Twitter
- Post-publication peer-reviews
- Facebook
- Sina Weibo
- Syllabi
- Wikipedia
- Google+
- LinkedIn
- Reddit
- Research highlight platform
- Q&A (Stack Overflow)
- Youtube
- Pinterest
- Patents





TEMA 4

# Tema 4. ¿Qué hago con mis datos?

1. Gestionar perfil
2. Responsabilidades académicas
3. La labor del investigador





TEMA 4

## SEIS pasos para incrementar la visibilidad de tu investigación



1. Gestionar **ORCID** para distinguirte tú y tu trabajo, del de todos los demás investigadores.
2. Comparte los resultados de tu investigación.
3. Gestiona y mantén actualizado el perfil en línea (o un CV web)
4. Participar en comunidades de redes sociales.
5. Blog
6. Tweet

Tomado de: <https://pitt.libguides.com/researchvisibility>



TEMA 5

# EJEMPLOS Y APLICACIONES

- Comportamiento del consumidor
- Campañas en salud
- Educación
- Tendencias varias disciplinas





TEMA 5

## EJEMPLOS Y APLICACIONES

- Global Pulse:
  - Ébola
  - Desastres naturales
- UN Global Pulse:
  - Migración y políticas públicas
- Twitter
  - Violencia, nutrición, cambio climático





TEMA 5

# PRÁCTICAS DIGITALES EMERGENTES

- Bots
- Feak news
- The Billion Prices Project MIT
- Toole





TEMA 5

# DISCIPLINAS QUE UTILIZAN BIG DATA

- Práctica política
- Periodismo
- Derecho





TEMA 5

# LOS DATOS COMO CONSTRUCCIÓN SOCIAL

Aprendizaje automatizado

Minería de Datos



# Bibliografía



Behavioral Experiments With Social Algorithms: An Information Theoretic Approach to Input–Output Conversions. (s. f.). Recuperado 24 de junio de 2019, de MartinHilbert.net website: <https://www.martinhilbert.net/behavioral-experiments-with-social-algorithms-an-information-theoretic-approach-to-input-output-conversions/>

Donoho, D. (2017). 50 Years of Data Science. *Journal of Computational and Graphical Statistics*, 26(4), 745-766. <https://doi.org/10.1080/10618600.2017.1384734>

Hox, J. J. (2019). *Computational Social Science Methodology, Anyone?* 10.

Rocha, M. E. M. (2018). Grandes datos, grandes desafíos para las ciencias sociales. *Revista Mexicana de Sociología*, 80(2). <https://doi.org/10.22201/iis.01882503p.2018.2.57723>

Rogel-Salazar, R. (2019, junio 15). Estrategias de visibilidad de la producción académica de las politólogas en medios digitales. <https://doi.org/10.6084/m9.figshare.8279720.v2>

Science Communication on Social Media | InformalScience.org. (s. f.). Recuperado 20 de junio de 2019, de <https://www.informalscience.org/news-views/science-communication-social-media>

Véliz, C. (2019, junio 14). Tribuna | Inteligencia artificial: ¿progreso o retroceso? El País. Recuperado de [https://elpais.com/elpais/2019/06/13/opinion/1560421249\\_824783.html](https://elpais.com/elpais/2019/06/13/opinion/1560421249_824783.html)

