



## ¿QUE ES *BIG DATA* ?

*Big Data*, macrodatos o datos masivos es un concepto que hace referencia al almacenamiento de grandes cantidades de datos y a los procedimientos usados para encontrar patrones repetitivos dentro de esos datos.

*Wikipedia, acceso: 14/01/2017*

“Aquel conjunto de datos que, por su tamaño ingente, sobrepasa la capacidad de ser gestionado por bases de datos de integración tradicionales”.

*José Manuel Martínez Sesmero, Farmacia hospitalaria*

La mayoría de profesionales consideran múltiples *terabytes* o *petabytes* como el punto de referencia del *Big Data*. Otros, dudan en comprometerse con un volumen de datos específico, ya que el desarrollo tecnológico conlleva a que el concepto actual de grande sea en el futuro, normal. Otros, definirán *Big Data* según el contexto. En otras palabras, *Big Data* es una etiqueta subjetiva vinculada a situaciones en las que las infraestructuras humanas y técnicas no pueden adaptarse a las necesidades de datos.

*Villanova University, Philadelphia*

Grandes volúmenes de datos generados a alta velocidad, complejos y diversos que requieren técnicas y tecnologías que permitan la captura, el almacenamiento, la distribución, la gestión y el análisis de la información.

*Report from the U.S. Congress, August 2012*

*IHTT: Transforming Health Care through Big Data Strategies for leveraging big data in the health care industry; 2013*



## LAS 3 Vs CLÁSICAS de *BIG DATA*:

### *Volumen:*

- ü Generación 2,9 trillones de bytes de datos / día (5.800 x 10<sup>6</sup> discos duros medios)
- ü 90% de los datos del mundo se han creado en los dos últimos años

### *Variedad:*

- ü Datos estructurados, semi-estructurados, no estructurados
- ü 80% de los datos son NO estructurados
- ü Diversidad datos no estructurados (texto, imágenes, ...)

### *Velocidad:*

- ü Vida útil de los datos
- ü Velocidad de gestión
- ü Rapidez en que los datos pueden analizarse

# VOLUMEN. ALMACENAJE DE DATOS EN EL SISTEMA DE SALUD, de E.U.A.

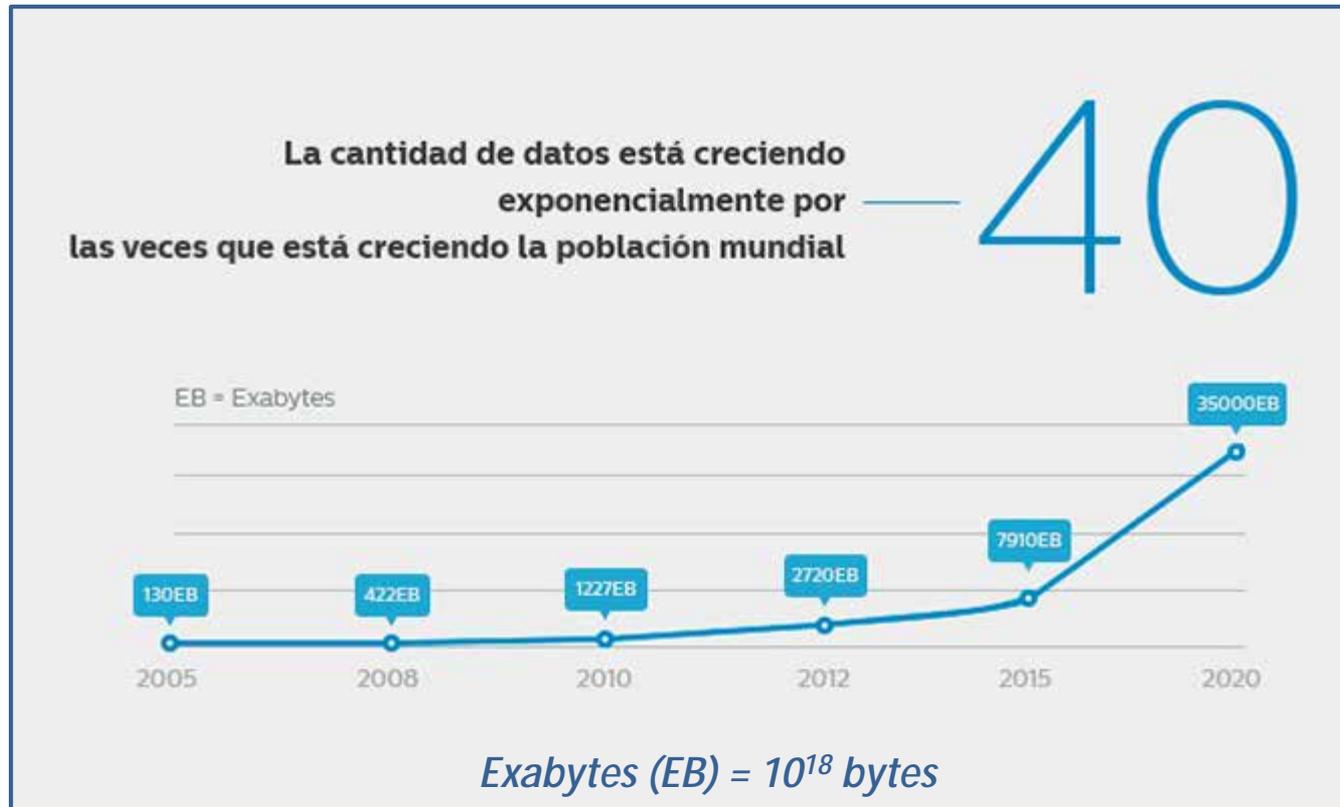


Figura: Martínez Sesmero JM. *Big data: aplicación y utilidad para el sistema sanitario*. Farm Hosp 2015; 39(2): 69-70.

## VOLUMEN. RELACIÓN ENTRE LAS UNIDADES DE ALMACENAMIENTO

NOMBRE	MEDIDA BINARIA	BYTES	EQUIVALENTE
Megabyte (MB)	$2^{20}$	1.048.576	1024 KB
Gigabyte (GB)	$2^{30}$	1.073.741.824	1024 MB
Terabyte (TB)	$2^{40}$	1.099.511.627.776	1024 GB
Petabyte (PB)	$2^{50}$	1.125.899.906.842.624	1024 TB
Exabyte (EB)	$2^{60}$	1.152.921.504.606.846.976	1024 PB
Zettabyte (ZB)	$2^{70}$	1.180.591.620.717.411.303.424	1024 EB
Yottabyte (YB)	$2^{80}$	1.208.925.819.614.629.174.706.176	1024 ZB

Cada letra, número o signo de puntuación ocupa un byte (8 bits). Un archivo de texto de 5.000 bytes equivale a 5.000 letras o caracteres.



# VELOCIDAD de la GENERACIÓN de los DATOS:



# ANÁLISIS EN TIEMPO REAL:



# ¿BIG DATA PUEDE TRANSFORMAR LA ASISTENCIA SANITARIA?

- 1 Apoyo a la investigación.** Genómica. Mejorar métodos diagnósticos, terapias, resultados. Reducir costes (*medicina personalizada*).
- 2 Transformar datos en información:** traducir datos no estructurados en estructurados. Validar e intercomunicar datos para mejorar la atención sanitaria de forma eficiente (*generación nuevo conocimiento*).
- 3 Apoyar el autocuidado de los pacientes:** tecnología móvil para educar y motivar cambios que mejoren la gestión de enfermedades y la toma de decisiones saludables (*salud comunitaria*).
- 4 Dar soporte a los proveedores :** aliviar presión en los centros sanitarios para ahorrar costes y mejorar resultados clínicos (*medicina predictiva*).
- 5 Aumentar la conciencia:** Apoyar la prevención y priorizar problemas de salud pública (*medicina preventiva*).
- 6 Pool data:** nuevos tipos de análisis que respondan a preguntas, algunas de las cuales todavía no se han planteado .

### *Columbia University Medical Center's, E.U.A.*

Detección de complicaciones severas en pacientes con hemorragia cerebral por ruptura de aneurisma 48h antes que con los procedimientos diagnósticos estándares.

### *Rizzoli Orthopedic Institute. Bologna, Italia.*

Identificación de pacientes de alto riesgo a partir de la severidad de los síntomas. Reducción anual del 30% de hospitalizaciones y un 60% de pruebas de imagen.

### *Hospital for Sick Children (Sick Kids). Toronto, Canadá.*

Identificación de signos vitales en la monitorización de bebés propensos a infección nosocomial unas 24 h antes que con los métodos estándares.

### *University of Michigan Health System, E.U.A.*

Experiencia profesional cruzada con la información de las bases de datos de transfusiones: reducción del 31% en transfusiones y de 200.000 dólares de costes / mes.

### *California-based Kaiser Permanente, E.U.A.*

Historia clínica > 10<sup>6</sup> de pacientes tratados con analgésico *Vioxx* determinó un riesgo 3 veces superior de padecer un infarto. *Vioxx* se retiró del mercado.

FDA estimó > 27.000 casos de infarto y muerte súbita provocados por el uso de *Vioxx*.

*IHTT: Transforming health Care through Big Data Strategies for leveraging big data in the health care industry, 2013 ; David J Graham et al. The Lancet 2005; 365 (9458): 475–481.*

### *Investigación Translacional*

Proyecto eMERGE: Combina biorepositorios de ADN con informes médicos electrónicos.

*Gottesman O. The electronic Medical Records and Genomics (eMERGE) network: past, present and future. Genet Med 2013. Chute CG. Some experiences and opportunities for big data in translational research. Genet Med 2013.*

### *Medicina comunitaria, New Zealand*

Procesar el lenguaje natural (texto) para estimar la prevalencia y la gravedad de las enfermedades a partir de la historia de salud electrónica de Nueva Zelanda.

*MacRae J. Accessing primary care big data: the development of a software algorithm to explore the rich content of consultation records. BMJ open 2015.*

### *Google en Salud Pública*

La tendencia de la gripe en google predice epidemias 1 semana antes que CDC.

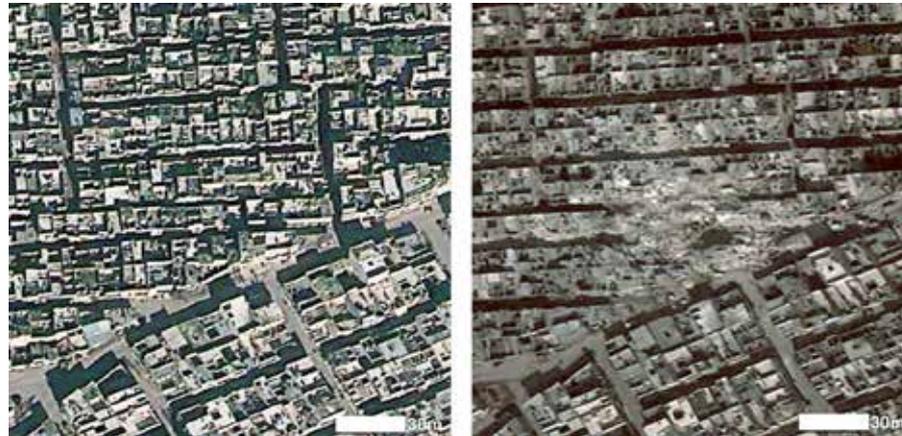
*Johns Hopkins School of Medicine, E.U.A. Google Flu Trends & Google Dengue Trends*  
(<https://www.google.org/flutrends/about/>)

### *"Peace Technologies"*

*Twitter:* el análisis de las actualizaciones de Twitter fue tan preciso como los informes del seguimiento de la propagación del cólera en Haití (terremoto del 2010).

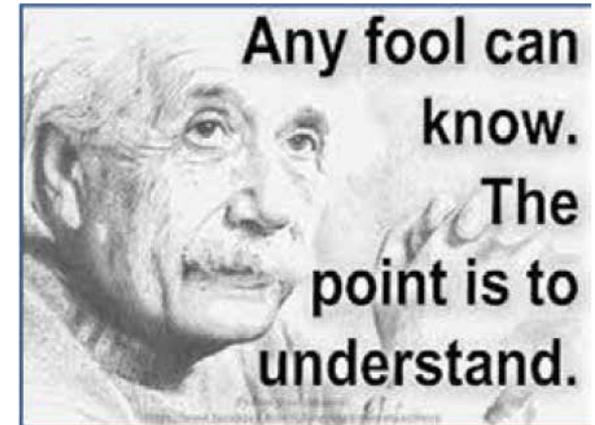
*Ushahidi:* nombre en swahili para "testimonio", creado por bloggers kenianos e informáticos en respuesta a la violencia relacionada con las elecciones en Kenia en 2007.

*Imágenes de satélite de alta resolución:* daños a edificios en la ciudad de Aleppo, Siria



## LIMITACIONES de *BIG DATA*:

- ü Metodología manejo de los datos
  - ü Información en silos (interoperabilidad limitada)
  - ü Valores perdidos
  - ü Calidad de la información, "*ruido de datos*"
  - ü Interpretación de los datos
  - ü Sesgo de selección
  - ü Problemas de dependencia entre variables
  - ü Diferenciar asociación respecto a causalidad
  - ü Correlaciones casuales (no causales !!)
  - ü Inferencia estadística sesgada
  - ü Incapacidad de integrar procesos de aprendizaje y mejora continua
- 
- ü *Anonimización de datos ¿Es posible?...*
  - ü *Legislación actual*



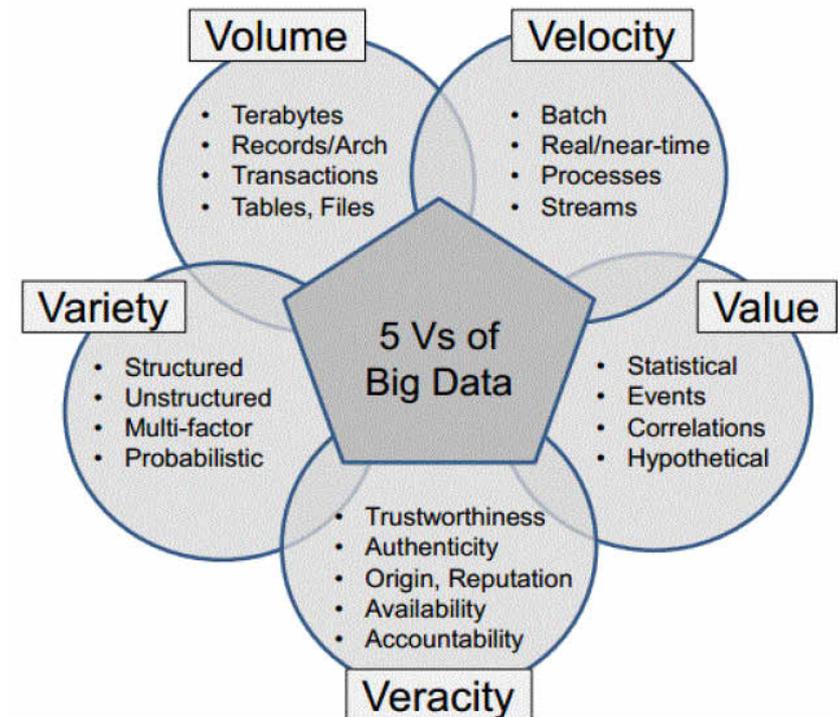
*Schneeweiss S. Learning from Big Health Care Data. N Engl J Med 2014; 370 (23): 2161-2163.*  
*Raghupathi W et al. Big Data analytics in healthcare: promise and potential. HISS 2014*

# LAS 5 Vs ACTUALES de *BIG DATA*:

## Volumen, Variedad, Velocidad y ...

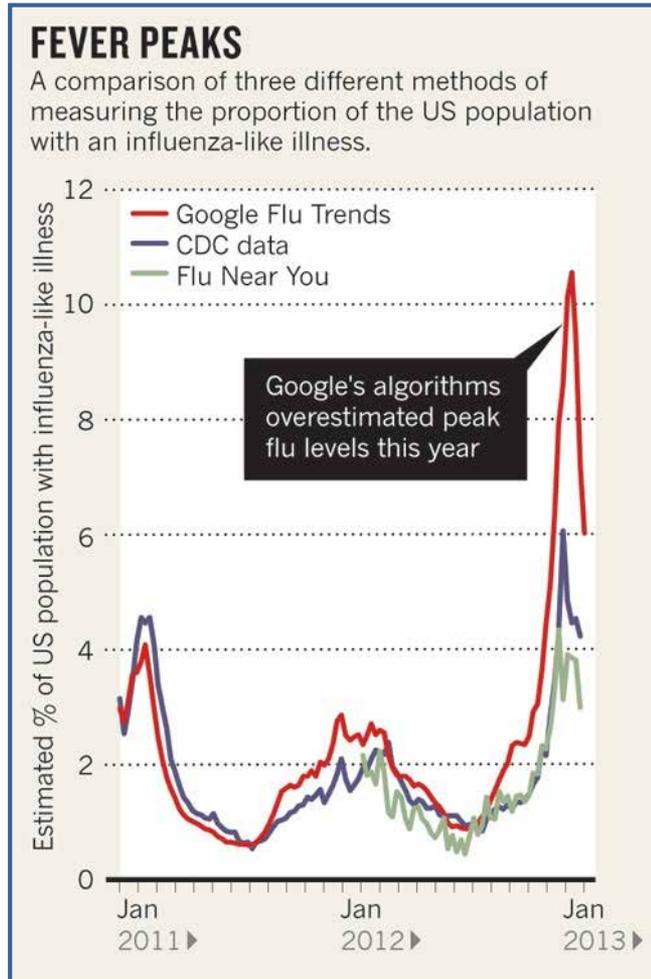
**Veracidad:** El gran objetivo (confianza); Calidad e integridad de los datos (fiabilidad)

**Valor:** Beneficio “real” económico y social de la aplicación del análisis de los datos.



# WHEN GOOGLE GOT FLU WRONG

US outbreak foxes a leading web-based method for tracking seasonal flu.



*Declan Butler. Nature 2013; 494: 155-6 .*

*Source: Google Flu Trends ([www.google.org/flutrends](http://www.google.org/flutrends)); CDC; Flu Near You*



16 octubre 2014

JAVIER LÓPEZ TAZÓN

*Peligro, sus datos están a la venta*



"...una de cal y otra de arena..."



Graffiti en La Ferté-sous-Jouarre,  
Seine-et-Marne, Francia.

Òscar Canalís Hernández

21 diciembre 2016

EI PAIS



Més valor a la informació  
de salut de Catalunya

## CATALUNYA: PROYECTO VISC + , AQUAS

El gobierno de Cataluña venderá los datos de los pacientes de la sanidad pública a empresas privadas *Cafè amb llet, Septiembre 2014*

La Generalitat planea vender datos sanitarios anónimos al sector privado. *El País, octubre 2014*

Salut rectifica i exclou els socis privats del nou arxiu amb dades dels pacients. *Ara.cat, enero 2015*

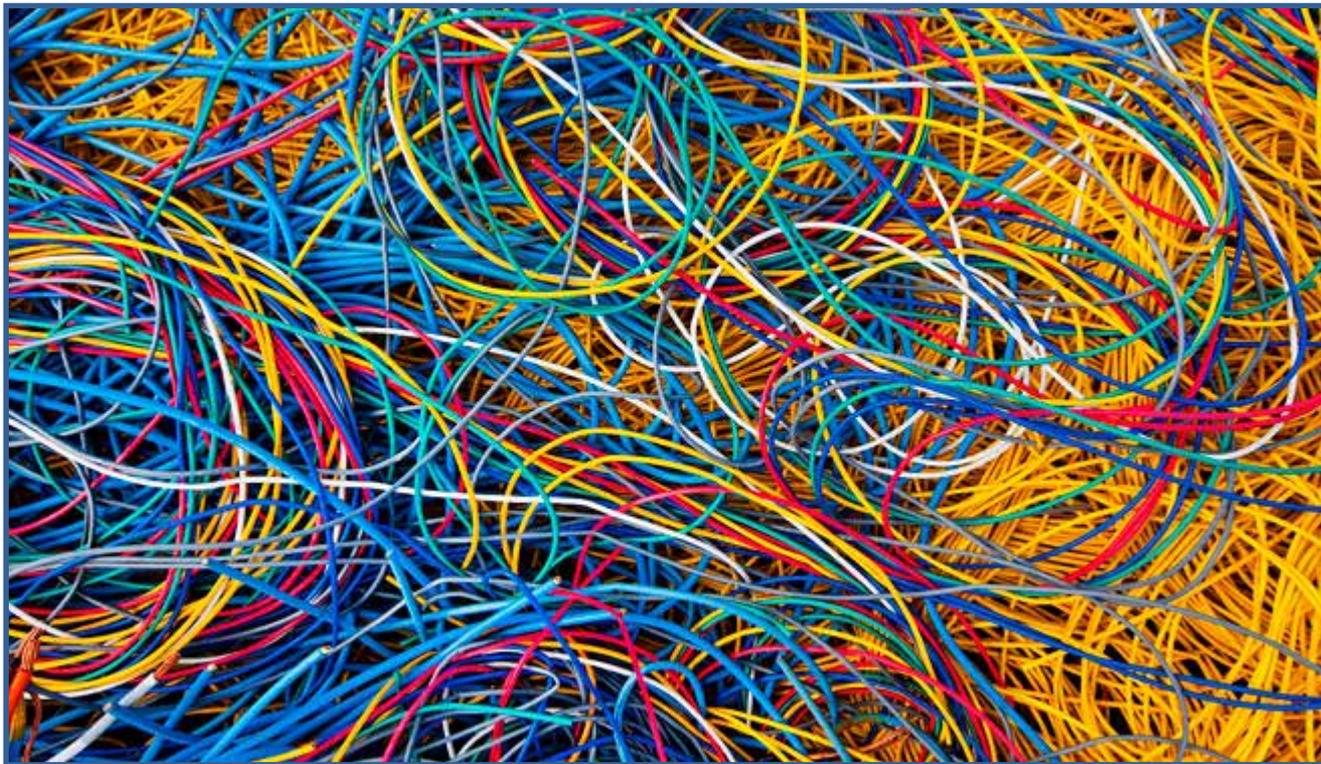
Científics catalans demanen a Salut que espavili i posi en marxa l'arxiu de dades sanitàries. *Ara, Junio 2015*

- *Observatori de Bioètica de la Universitat de Barcelona (Enero, 2015)*
- *COMB (abril, 2015)*
- *Síndic de Greuges de Catalunya (junio 2015)*

Salut se reune con los grupos parlamentarios para analizar el proyecto VISC+. *Departament Salut, nota de prensa. Enero 2016*

Salut rebateja el projecte per cedir dades sanitàries als investigadors (PADRIS). *Ara, Enero 2017*

# EXPERIENCIA EN CATALUNYA



PLAN DIRECTOR DE ONCOLOGIA

# FUENTES DE INFORMACIÓN PARA LOS INDICADORES DE CÁNCER DE CATALUNYA

**GLOBOCAN 2012**  
ESTIMATED CANCER INCIDENCE, MORTALITY AND PREVALENCE WORLDWIDE IN 2012

**World Health Organization**  
International Agency for Research on Cancer

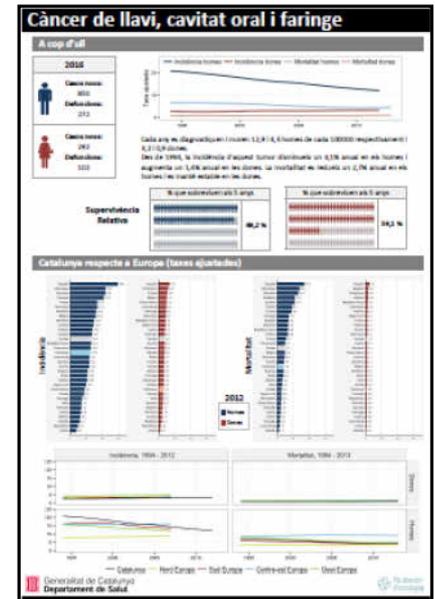
**Cancer Incidence in Five Continents Vol. VIII**  
**Cancer Incidence in Five Continents Vol. X**  
**Cancer Incidence in Five Continents Vol. IX**

**WHO**  
Cancer Incidence Database

**EUROCORE**  
Survival of cancer patients in Europe

Ist. Naz. Tumori, Milano & Ist. Superiore di Sanità

INTERNACIONAL

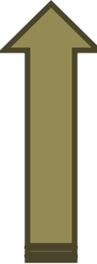


Registro Cáncer Girona

Registro Cáncer Tarragona

Registro Mortalidad Catalunya

Población catalana IDESCAT



CATALUNYA

# CÁNCER DE LABIO, CAVIDAD ORAL Y FARINGE. CATALUNYA 1994-2016

**2016**



Casos nous:

850

Defuncions:

272

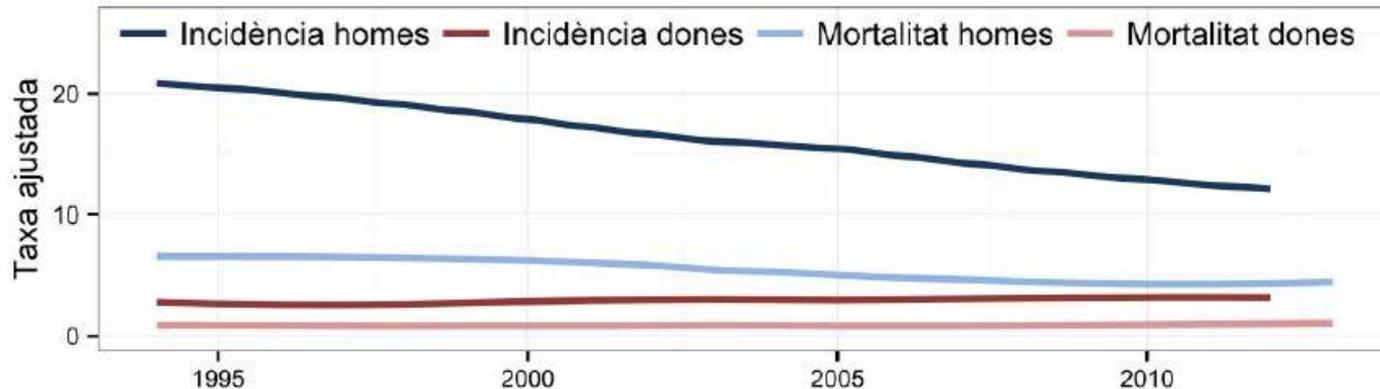


Casos nous:

292

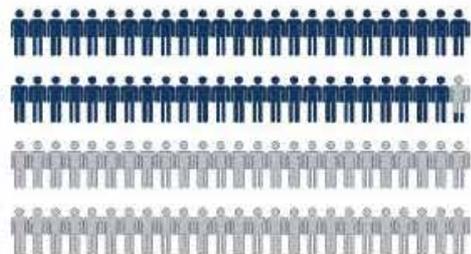
Defuncions:

102



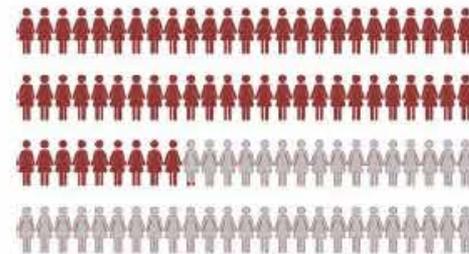
**Supervivència  
Relativa**

% que sobreviuen als 5 anys



49,2 %

% que sobreviuen als 5 anys

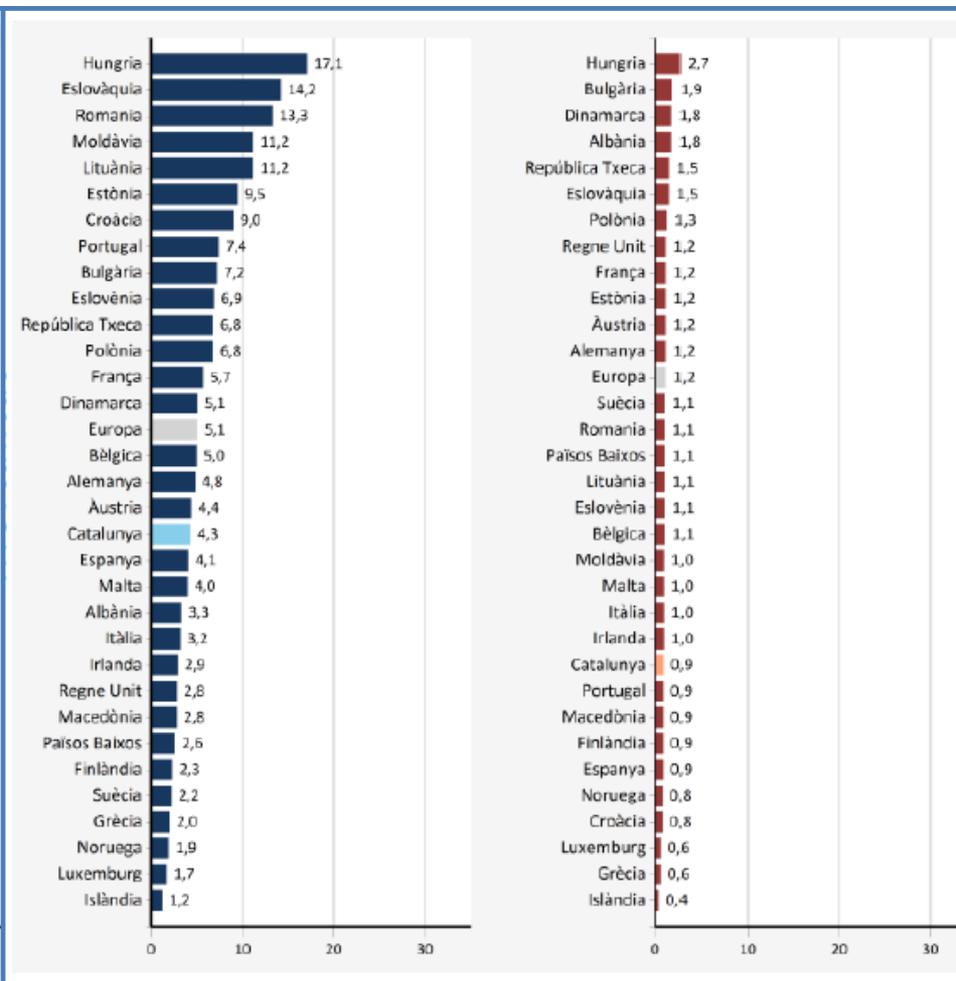
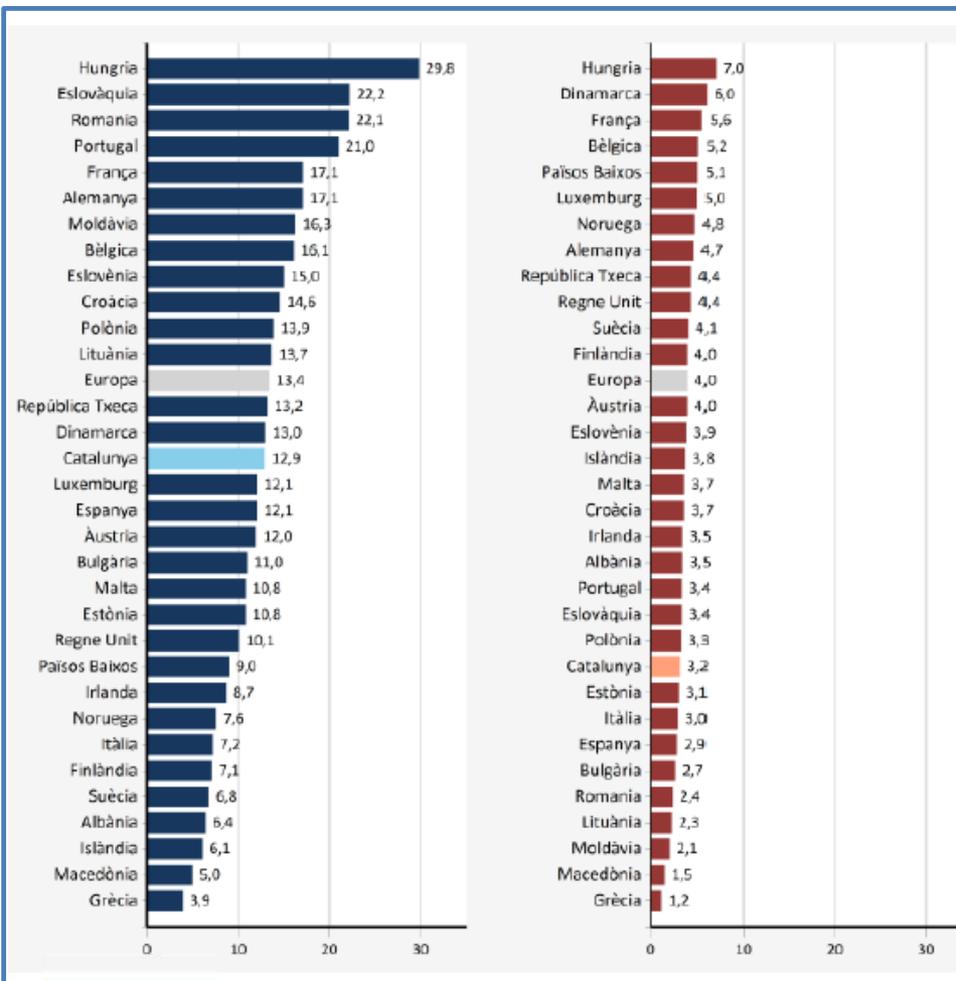


59,1 %

# CÁNCER DE LABIO, CAVIDAD ORAL Y FARINGE. CATALUNYA 2012

## INCIDENCIA

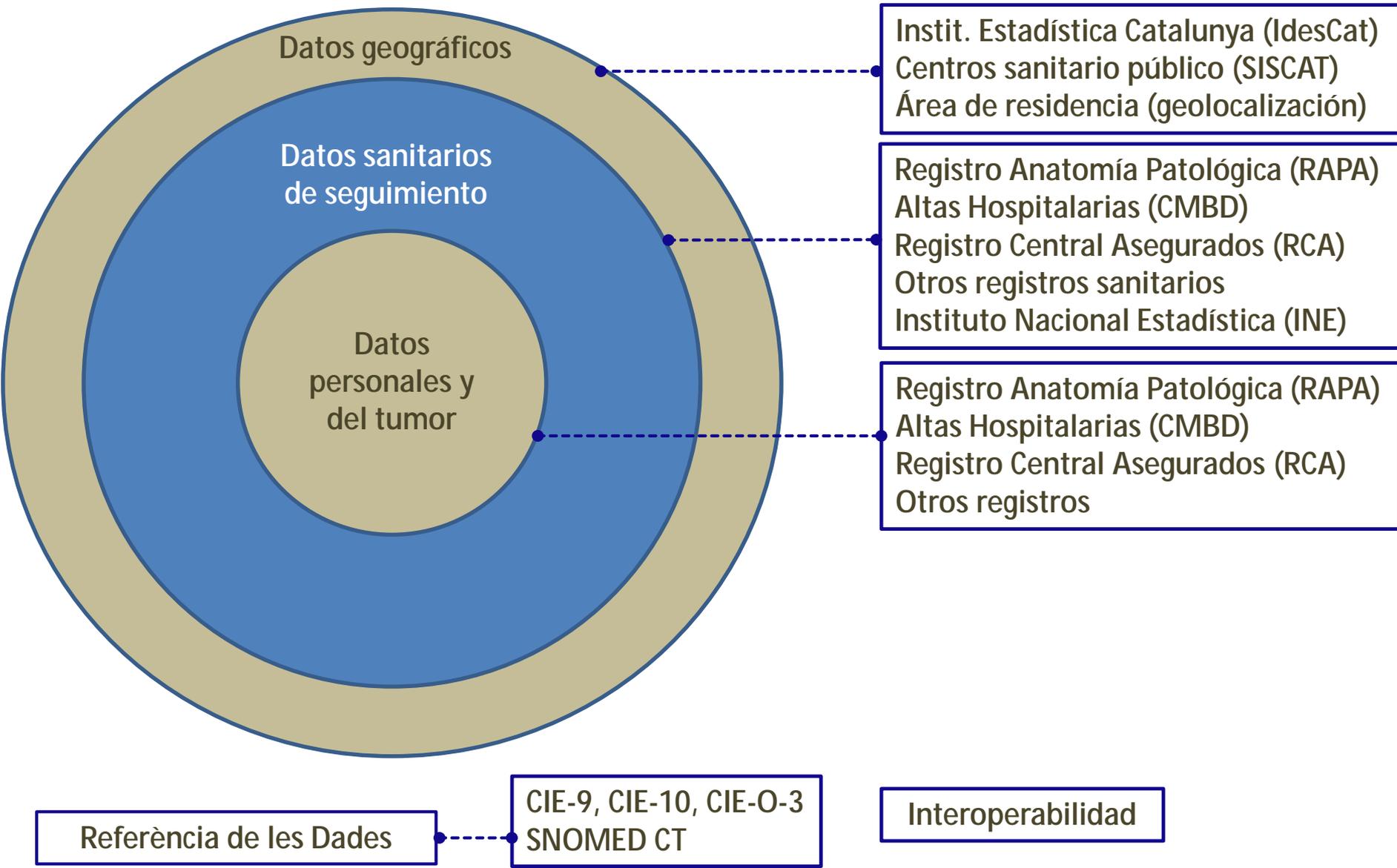
## MORTALIDAD



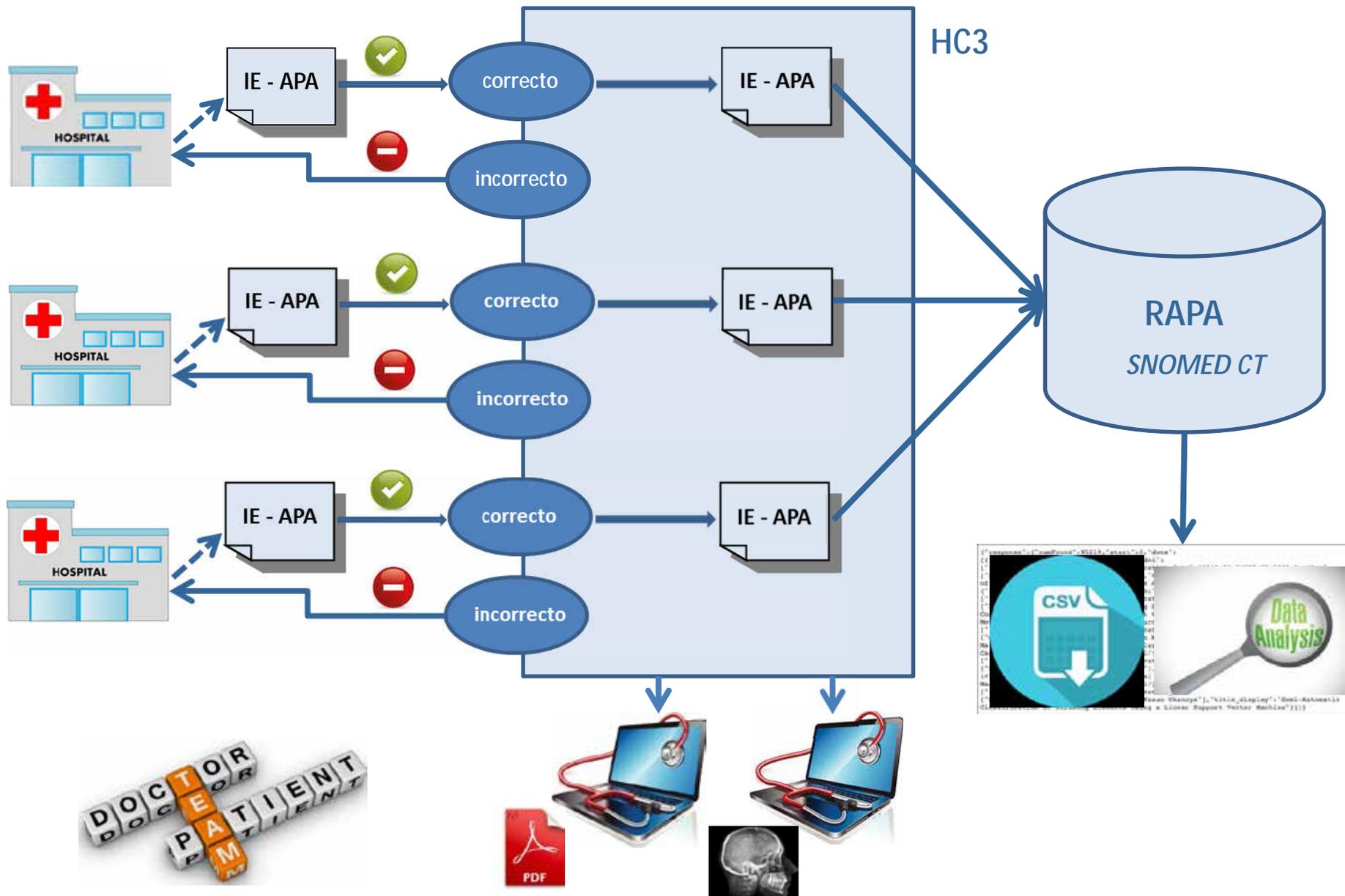
■ Homes

■ Dones

# INFORMACIÓN INTEGRADA SOBRE EL CÁNCER

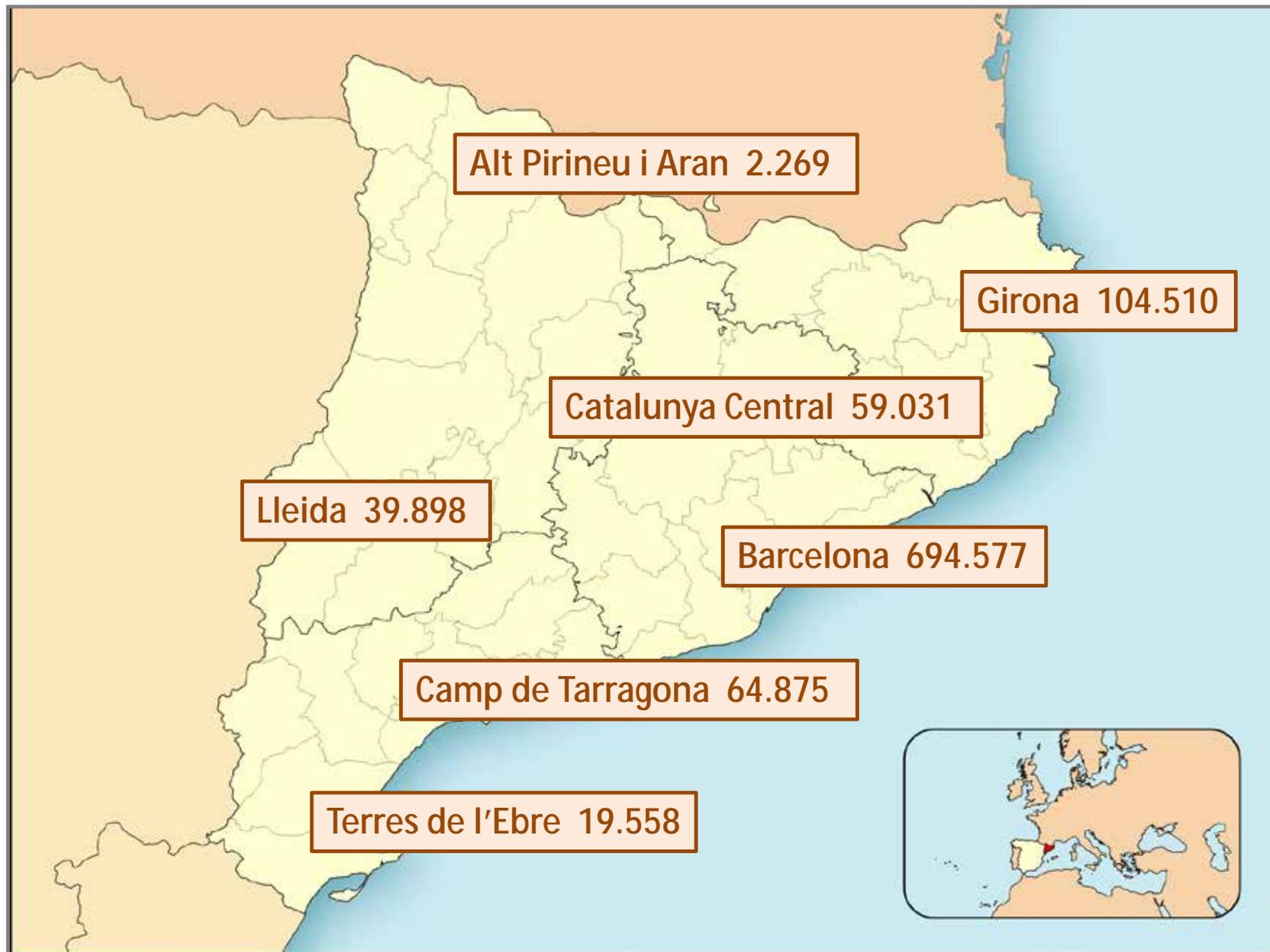


# NUEVA FUENTE DE INFORMACIÓN: REGISTRO DE ANATOMÍA PATOLÓGICA

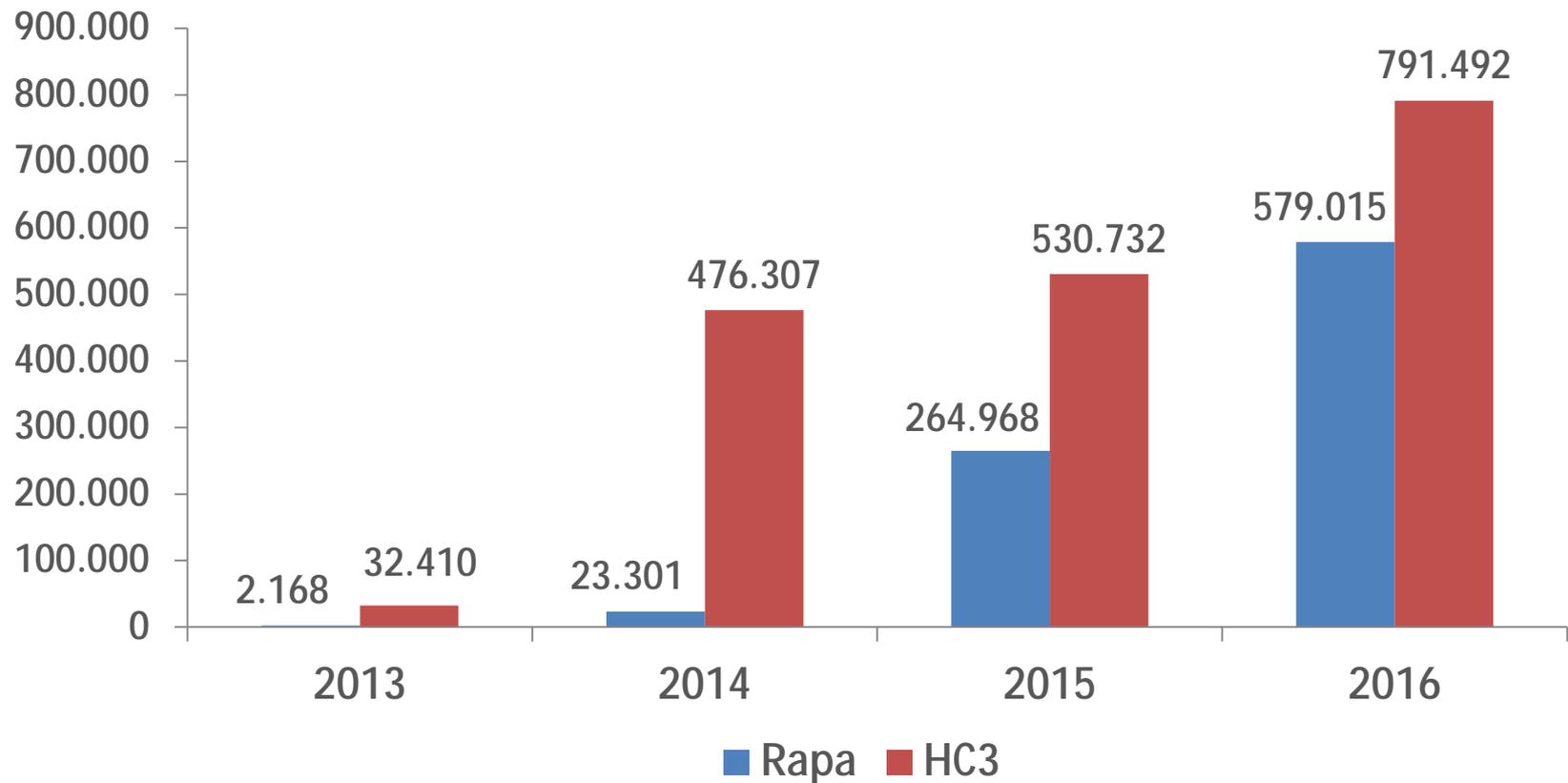


IE-APA: informe electrónico de Anatomía Patológica; HC3: Historia Clínica Compartida; RAPA: Registro Anatomía Patológica

# ANATOMIA PATOLÓGICA ESTRUCTURADA (APAE): ENERO-NOVIEMBRE 2016



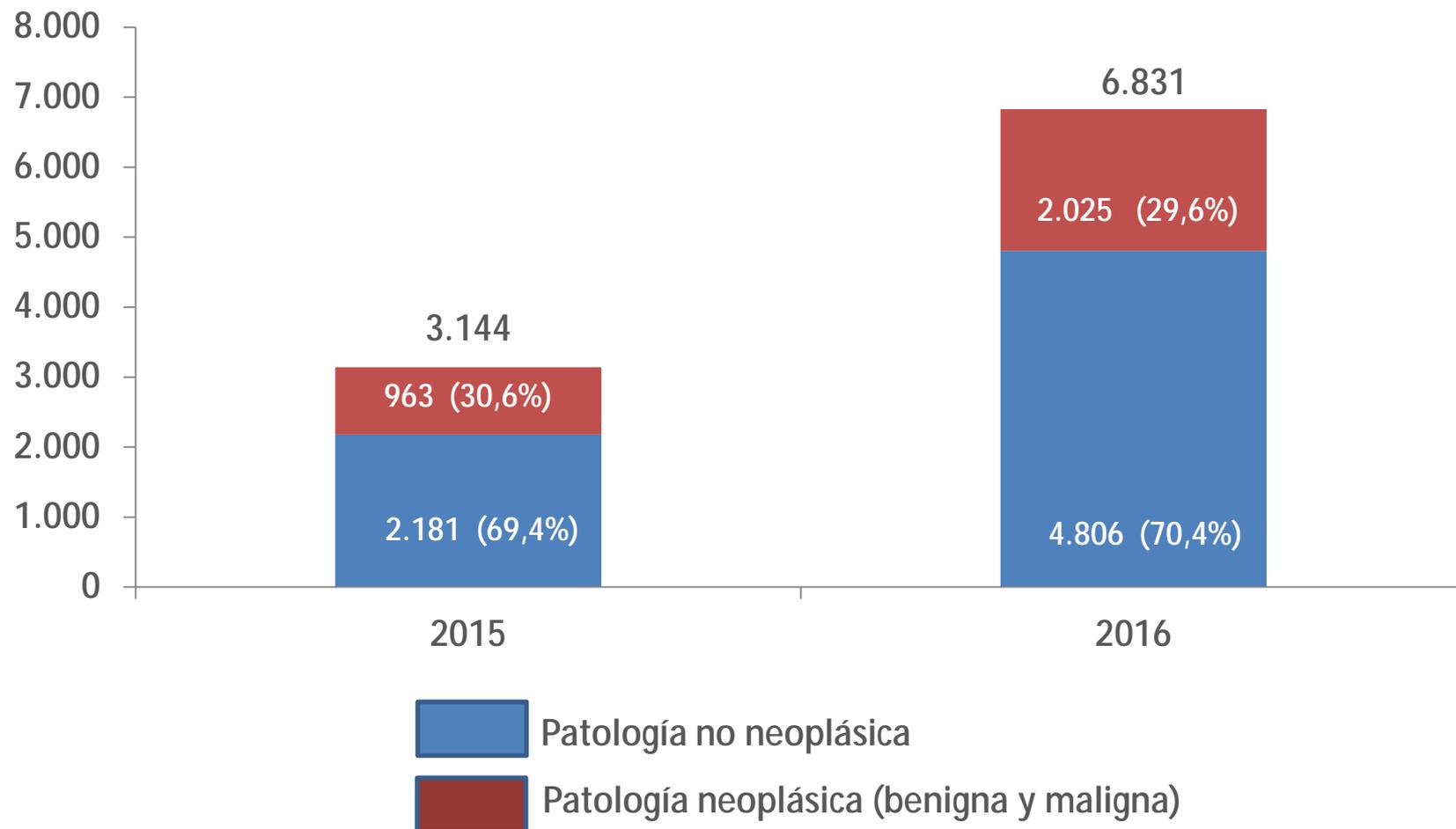
# RAPA. Informes Anatomía Patológica / año Catalunya, 2015 - 2016



Casparie M et al. *Pathology databanking and biobanking in The Netherlands, a central role for PALGA, the nationwide histopathology and cytopathology data network and archive*. Cellular Oncology 29 (2007): 19-24.

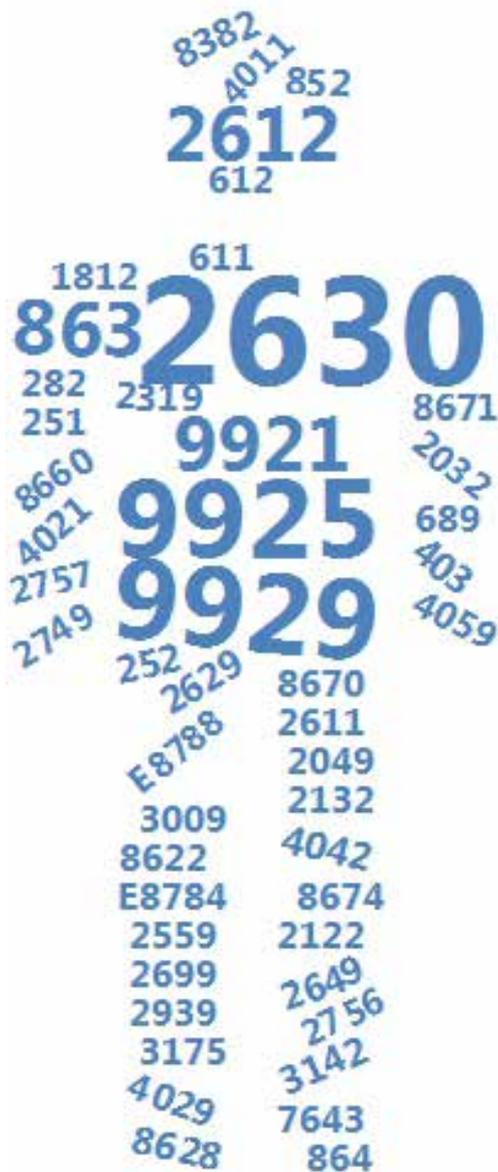
# RAPA. PERSONAS con patología en labio, cavidad oral y faringe

## Catalunya, 2015-2016





# CÁNCER DE GLÁNDULAS SALIVALES. CATALUNYA 2015



N= 136

Hombres	79	(58.1)
Morfología		
Sin APA	101	(74.3)
Carcinoma escamoso	16	(11.8)
Adenocarcinoma	3	(2.2)
Carcinoma	11	(8.1)
Linfoma	3	(2.2)
Tratamiento		
Quimioterapia	25	(18.4)
Radioterapia	2	(1.5)
Cirugía	106	(77.9)
Cirugía maxilofacial <sup>1</sup>	100	(73.5)
<i>Sialoadenectomía total (26.32)</i>	40	
<i>Sialoadenectomía parcial (26.31)</i>	39	
<i>Sialoadenectomía NE (26.30)</i>	23	
Nivel complejidad hospital <sup>2</sup>		
Alta tecnología	42	(36.5)
Alta resolución/referencia	61	(53.0)
Servicio del alta <sup>2</sup>		
Cirugía maxilofacial	47	(40.9)
Otorrinolaringología	59	(51.3)
Cirugía plástica	5	(4.3)

(1) Procedimientos ICD-9: 22.X, 25.X-29.X, 76.X

(2) % sobre procedimientos de cirugía maxilofacial

# CÁNCER DE CAVIDAD ORAL. CATALUNYA 2015



N= 209

Hombres	117	(56.0)
Morfología		
Sin APA	140	(67.0)
Carcinoma escamoso	58	(27.8)
Adenocarcinoma	1	(0.5)
Carcinoma	7	(3.3)
Linfoma	2	(1.0)
Tratamiento		
Quimioterapia	48	(23.0)
Radioterapia	6	( 2.9)
Cirugía	169	(80.9)
Cirugía maxilofacial <sup>1</sup>	160	(76.6)
<i>Escisión cavidad oral (27.49)</i>	99	
<i>Mandibulectomía parcial (76.31)</i>	49	
<i>Reparación plástica cavidad oral (27.57)</i>	41	
<i>Ostectomía parcial hueso facial (76.39)</i>	23	
<i>Reconstrucción mandíbula (76.43)</i>	18	
<i>Odontectomía (23.19)</i>	17	
Nivel complejidad hospital <sup>2</sup>		
Alta tecnología	215	(66.0)
Alta resolución/referencia	97	(29.8)
Servicio del alta <sup>2</sup>		
Cirugía maxilofacial	261	(80.1)
Otorrinolaringología	50	(15.3)
Cirugía plástica	11	( 3.4)

(1) Procedimientos ICD-9: 22.X, 25.X-29.X, 76.X

(2) % sobre procedimientos de cirugía maxilofacial

# CÁNCER DE FARINGE. CATALUNYA 2015



N= 326

Hombres	248	(76.1)
Morfología		
Sin APA	227	(69.6)
Carcinoma escamoso	64	(19.6)
Adenocarcinoma	4	(1.2)
Carcinoma	17	(5.2)
Linfoma	14	(4.3)
Tratamiento		
Quimioterapia	163	(50.0)
Radioterapia	23	(7.1)
Cirugía	137	(42.0)
Cirugía maxilofacial <sup>1</sup>	76	(23.3)
Amigdalotomía (28.2)	32	
Gastrotomía percutánea (43.11)	26	
Resección tejido de la faringe (29.39)	22	
Resección tejido de laringe (30.09)	19	
Faringectomía parcial (29.33)	17	
Inserción sonda (96.04; 96.07)	19	
Nivel complejidad hospital <sup>2</sup>		
Alta tecnología	67	(52.3)
Alta resolución/referencia	56	(43.8)
Servicio del alta <sup>2</sup>		
Cirugía maxilofacial	16	(12.5)
Otorrinolaringología	96	(75.0)
Cirugía plástica	13	(10.2)

(1) Procedimientos ICD-9: 22.X, 25.X-29.X, 76.X

(2) % sobre procedimientos de cirugía maxilofacial

## CONCLUSIÓN:

La **primera revolución** de la tecnología de la información en medicina fue la digitalización del historial médico.

La **segunda revolución** será aprovechar la información contenida en el historial médico digitalizado y combinarlo con otras fuentes para mejorar la asistencia sanitaria.



**¡ GRACIAS !**

