



Islas Canarias
Del 15 al 19 de noviembre de 2021



Instituto de Estadística
y Cartografía de Andalucía
Consejería de Transformación Económica,
Industria, Conocimiento y Universidades

DIFUSIÓN DE LAS ESTADÍSTICAS DEL COVID

Autor/a 1

Jose Antonio Moreno
josea.moreno.munoz@juntadeandalucia.es

Autor/a 2

Esther Barbero Areales
esther.barbero@juntadeandalucia.es

PALABRAS CLAVE

COVID / reutilización / interoperabilidad / salud



Índice

1. Introducción.....	3
2. Objetivos.....	3
3. Metodología.....	4
3.1. Y recién estrenado el estado de alarma nos ponemos las pilas.....	4
3.2. Avance de los servicios interoperables gracias al COVID	6
3.3. Proceso de obtención del dato por parte de la Consejería de Salud	7
4. Resultados.....	10
5. Conclusiones.....	13
6. Principales referencias bibliográficas	13

1. Introducción

El 14 de marzo de 2020 se decreta el estado de alarma en todo el territorio nacional, para afrontar la situación de emergencia sanitaria provocada por el coronavirus COVID-19. Este decreto promueve medidas de distanciamiento social y solo permite desplazamientos de primera necesidad. Es a partir de esta fecha cuando se inicia el confinamiento y gran parte de los trabajadores en toda España, sobretodo en el sector servicios, empiezan a teletrabajar.

Es también a partir de este momento cuando la población empieza a tomar conciencia de la gravedad del asunto y comienza la alarma social, a ello se une el desconocimiento de la enfermedad, de su origen, cómo se transmite, cómo curarla, al mismo tiempo la falta de datos oficiales de población afectada, número de fallecidos, ingresados,... aunque se sabe que las cifras crecen de forma exponencial, pero en un primer momento se carece de datos oficiales.

Por tanto, la Consejería de Salud y Familias de la Junta de Andalucía, basándose en la experiencia previa de años difundiendo sus estadísticas a través del banco de datos del IECA (Instituto de Estadística y Cartografía de Andalucía), elabora un documento de propuesta de colaboración con dicho organismo, para hacer accesibles los datos sobre la evolución y comportamiento de la pandemia de forma visual y de manera que sean reutilizables por parte de los usuarios.

2. Objetivos

La propuesta contempla la elaboración de una página web que incluya un cuadro de mando con los principales indicadores, gráficos y mapas, que permitan de forma muy visual conocer la evolución de los indicadores y ofrezca la posibilidad de descargar los datos para su uso por investigadores, medios de comunicación y usuarios en general.

Se propone para su difusión la creación de un Portal específico para publicar todas las cifras estadísticas relacionadas con el COVID provenientes de la Consejería de Salud y Familias, de forma diaria, para los principales ámbitos territoriales (distritos sanitarios, provincias y Andalucía) y en diferentes formatos que hagan más amigable la interpretación y seguimiento de la evolución de la pandemia.

Un elemento clave y que fue objetivo prioritario desde el inicio del estado de alarma, fue incorporar toda la información a BADEA (Banco de Datos Estadístico de Andalucía), creado por el IECA y que es el repositorio de datos del SECA (Sistema Estadístico y Cartográfico de Andalucía). Dicho repositorio garantiza la unicidad del dato, que es almacenado en el sistema por el órgano que lo produce y además, dispone de las herramientas necesarias para que pueda ser difundido a través de diferentes formatos. Otro aspecto clave es que se facilita a los productores de información estadística “compartir” ese dato único de forma que se puedan combinar datos procedentes de diferentes actividades sin necesidad de “repetir” información en el sistema. En este sentido, una de las cifras más demandadas, las tasas a 7 y 14 días, combina información almacenada en este portal específico con la información procedente del padrón municipal de habitantes.

3. Metodología

3.1. Y recién estrenado el estado de alarma nos ponemos las pilas

Justo en los primeros días del confinamiento, la Consejería de Salud y Familias solicita al IECA colaboración para difundir la información relacionada con el COVID. Desde casa, muertos de miedo pero con unas ganas inusitadas por ayudar en lo que fuese necesario, el equipo de administradores de BADEA en coordinación directa con los técnicos de la consejería plantean la necesidad de generar una nueva actividad, llamada ‘Informe COVID-19 en Andalucía’, en la que se recogieran todas las cifras estadísticas que se fueran generando a partir de ese momento.

Al principio la novedad del teletrabajo conlleva cierta dificultad para trabajar en equipo desde la distancia, unido a las prisas por difundir ‘algo’, debido a lo complicado de la situación, en la que la falta de información oficial sobre la evolución y la situación de la enfermedad era cada vez más palpable y la incertidumbre sobre la información con la que se va a contar, la desagregación territorial con la que se van a cargar los datos y hasta qué nivel se va a permitir su difusión para respetar los criterios de secreto estadístico, la periodicidad con la que se va a cargar la información y también la periodicidad con la que se va a difundir, el plan de tabulación, etc., genera un pequeño (GRAN) caos, que tiene como primeras consecuencias las rectificaciones en la forma de organizar y estructurar la información en BADEA, llevando en algunas ocasiones a replantear completamente el trabajo realizado.

El trabajo en equipo entre el Servicio de Estadística de la Consejería de Salud y Familias, el Servicio de difusión del IECA y los administradores de BADEA era fundamental para poder dar una solución rápida, a medida que iban surgiendo nuevas necesidades de información y se iba definiendo la actividad.

En un primer momento una de las cuestiones que más costó concretar fue la periodicidad que iban a tener los cubos y aunque se tenía claro que la información se difundiría diariamente, la temporalidad mínima que ofrecía BADEA era el mes. Se tuvo que investigar la forma a través de la cual podíamos incorporar en BADEA el día como dimensión temporal pero tuvimos que descartar esa idea por los cambios que se habían de acometer y el impacto que tendría en cuanto a horas de dedicación. Y tras muchas vueltas, se optó por definir cubos anuales en los que se añadieron otras dimensiones que sin tener las características propias de dimensiones temporales, sí que ofrecían información del mes, la semana y el día.

Una vez resuelto el problema de la periodicidad y dando por hecho que se van a publicar los datos diariamente, se plantea el reparto de tareas para estar coordinados y garantizar que el proceso sea lo más ágil posible. Partiendo de la base de que una vez llega el dato al servicio de estadística de la Consejería de Salud y Familias se depura y se trata, hay que determinar quién carga el dato, quién lo publica, qué soporte de difusión usamos, en qué momento se difunde, etc. Y más pronto que tarde y a base de horas de dedicación y de muchas reuniones en horario indeterminado (alguna reunión se inició en un día y acabo en el siguiente...), se fueron concretando las distintas cuestiones. El Servicio de Gestión de la Información da soporte en todo lo relacionado con la generación de cubos y el diseño de la estructura de los datos para adaptarlo a BADEA siguiendo también los requisitos impuestos por los compañeros de la Consejería; también da soporte en el diseño de las consultas que conforman el plan de tabulación y difusión; la Consejería se encarga de cargar los datos, validarlos y confeccionar las consultas que se difunden a través del portal Web así como las consultas que van a alimentar el cuadro de mando diseñado en Tableau;

el Servicio de Difusión se encarga del diseño de ese cuadro de mando así como de dar el OK a la difusión diaria de los datos cargados; el Servicio de Informática se encarga de desarrollar los conectores para que Tableau pueda interoperar con los ficheros JSON que BADEA genera; etc.

Y ni que decir tiene que muchas cuestiones que iban surgiendo se iban respondiendo sobre la marcha y el correo electrónico, el teléfono y, sobre todo, el WhatsApp, eran herramientas de uso cotidiano que estaban activas las 24h del día. Era una cadena o secuencia de acciones que no teníamos prevista y a base de entrega y de un entusiasmo inusitado por todos los que participábamos en el proyecto, hoy lo podemos contar como un caso de éxito.

Otro punto de debate fue la dimensión territorial. Desde el principio se tuvo claro que la información se iba a cargar a nivel municipal, pero solo se le daría difusión a nivel provincial así como para el total de Andalucía y los distritos sanitarios, pero cada vez había una presión mayor para que se ofreciese información municipal, que era lo que realmente demandaban más los usuarios, ya que la enfermedad iba evolucionando de forma diferente en función del territorio y se hacía necesario saber cómo estaba la situación a nivel municipal. Inicialmente había muchas reticencias por parte de la Consejería a dar este nivel de detalle, porque se pensaba que se podría estar vulnerando el secreto estadístico y tal vez también, porque se desconfiaba de la fiabilidad de los datos a ese nivel de detalle, con los problemas que eso podría suponer, sobretodo, en municipios pequeños, y finalmente fue tal la presión, que hubo que facilitar información municipal. Y sería justamente esta información la que con el paso del tiempo se convertiría en el referente para que el comité de expertos la tuviese como referente de cara a la imposición de restricciones en función de las tasas municipales de casos confirmados en los últimos 14 días por cada 1.000 habitantes.

En cuanto al territorio, generó también alguna confusión el hecho de que se publicasen los datos como requería la Consejería de Salud, por provincias sanitarias y no por provincias geográficas. Una provincia sanitaria está compuesta por sus distritos sanitarios, y estos pueden contener zonas básicas de salud de municipios limítrofes que corresponden geográficamente a una provincia diferente, por ejemplo, el caso de Peñafior, que siendo un municipio de Sevilla está incluido en el distrito sanitario Guadalquivir, que es de la provincia de Córdoba, o algunos municipios de la provincia de Huelva como Hinojos, Arroyo Molinos de León o Zufre, que desde el punto de vista sanitario están en distritos sanitarios de la provincia de Sevilla. Esto originó numerosas llamadas de ayuntamientos y usuarios indicando que los municipios no estaban publicados en la provincia correcta e incluso que no los podían encontrar pero la Consejería fue bastante reacia a cambiar el modo en el que se difundía la información, por ser el distrito sanitario su ámbito de referencia, y finalmente los usuarios comenzaron a familiarizarse con los conceptos de Distrito Sanitario y Zonas Básicas de Salud.

3.2. Avance de los servicios interoperables gracias al COVID

Meses antes de la publicación del COVID, el IECA había empezado a trabajar en la nueva estrategia de difusión basada en la visualización de datos con el software de business intelligence TABLEAU, con lo que el propio IECA se convierte en reutilizador de sus datos. Este nuevo perfil ha incidido, sin lugar a dudas, en la mejora de los mecanismos de difusión, pero es a raíz de la publicación del COVID, cuando se avanza en mejorar los procedimientos de reutilización y evolucionar el servicio de datos interoperables API RestFul del Banco de Datos Estadísticos de Andalucía (BADEA).

Y fue el servicio de difusión del IECA quien empezó a detectar algunas de las deficiencias que tenía el Servicio Interoperable de datos de BADEA, las cuales se detectaron y corrigieron para poder poner en producción el Portal IECA sobre el COVID-19. Algunas de ellas fueron:

- El servicio de datos de BADEA, en sus primeras versiones, se traía los datos de la consulta tanto si los datos estaban en el estado aprobado técnico como si estaban en el estado publicado. Para no cometer el error de publicar datos en las visualizaciones que no estuviesen publicados o facilitar a los reutilizadores información no publicada, se mejoró el servicio de datos para que solo devolviese datos publicados.
- También se detectaron casos en los que el servicio de datos generaba un fichero JSON sin datos, por ejemplo en determinadas consultas con decimales, también en determinadas consultas que tenían fórmulas con tasas de variación o cuando en una medida se utilizaban caracteres no habituales, como podía ser el uso del símbolo “<” (Incidencia acumulada en < 14 días), estos casos se fueron arreglando a medida que se detectaban.
- Otra mejora que se incorporó de cara a ofrecer una reutilización óptima en este caso a los reutilizadores no internos fue la incorporación a las consultas de BADEA de un icono que ofrece la url del servicio de datos asociado a dicha consulta, tanto de la consulta original, como de la consulta una vez filtrada, para obtener por ejemplo series temporales filtrando varios años.

Por todo esto, consideramos que la difusión de las estadísticas del COVID supuso el pistoletazo de salida en el ámbito de la interoperabilidad, algo en lo que se venía trabajando desde hacía tiempo, y que se vio materializado en ese momento gracias a la gran demanda de esta información por parte de organismos, prensa, universidades...

A modo de ejemplo de reutilizadores, podemos citar, al Instituto de Matemáticas de la Universidad de Sevilla que nos empezó a demandar información en formato CSV asociado a una url para disponer de información diaria y continua a partir de la cual generar sus propios estudios.

Y también nos encontramos dentro de la Junta de Andalucía, con la Consejería de Presidencia, Administración Pública e Interior, que ofrecía un resumen muy básico con información que diariamente se descargaban de nuestra página Web para actualizar la suya, porque desconocían que habíamos implementado los servicios de datos, que permitían hacer esto mismo de forma automática. Así, a mediados de abril de 2020, empezaron a consumir la información a partir de ficheros JSON que BADEA genera.

3.3. Proceso de obtención del dato por parte de la Consejería de Salud

Origen de los datos

Son los profesionales sanitarios de los Distritos Sanitarios y de los hospitales públicos de Andalucía, los encargados de grabar los casos confirmados en la aplicación Red Alerta del Servicio de Vigilancia y Salud Laboral de la Consejería de Salud y Familias. En el caso de los confirmados en hospitales privados, en un primer momento remitían los datos a las Delegaciones territoriales para su grabación en la aplicación Red Alerta pero dado el volumen de nuevos casos a grabar se desarrollaron herramientas de importación de resultados de laboratorio para poder incorporar los datos de forma masiva (actualmente los centros privados tienen acceso a Red Alerta).

También se incorporaban de forma masiva los datos clínicos de los casos, que se obtenían mediante cruce con ficheros procedentes del SAS y que al proceder de una fuente externa también requerían revisión y depuración.

Dado el volumen de casos se desarrolló un importador/actualizador automático que ha servido para importar resultados de laboratorios públicos y privados y una posterior actualización a partir de los datos generados por la aplicación de Historia utilizada por los rastreadores. Este importador también se ha utilizado para la mejora de calidad de datos demográficos, defunciones, fechas de alta hospitalaria, etc. Los datos se exportaban desde Red Alerta en un primer momento a ficheros Excel que se trataban con Access para estructurar la información y codificarla, de manera que pudiera ser leída por BADEA, organizándola en lo que llamamos cubos. Más adelante se desarrollaron una serie de scripts en R para obtener dichos cubos pero los archivos Excel empezaban a ser enormes, por lo que finalmente se optó por ajustar esos scripts para extraer la información de forma diaria y trabajarla en Oracle. Fue una adaptación continua a las circunstancias de cada momento.

La aplicación diseñada para la vigilancia epidemiológica no estaba pensada para soportar tal cantidad de registros (actualmente son alrededor de un millón de registros de los cuales el 80% corresponden a casos de Covid).

En periodos de aumento exponencial del volumen de casos se hizo necesaria la incorporación de grabadores (llegando a haber cientos de personas grabando datos de forma simultánea), muchas de ellas sin experiencia previa, lo que hizo imprescindible continuos controles de calidad para depurar duplicados y datos erróneos.

Y a todo ello, se unió que al tratarse de una situación tan vulnerable, se producían continuos cambios de criterio por parte del Ministerio, aparte de las continuas actualizaciones de protocolos, lo que provocaba estar continuamente adaptando los campos existentes en la aplicación y las definiciones.

Todo esto ha funcionado gracias al apoyo directo de la dirección (IECA, la DG Salud Pública y del Sv. Estadística de la Consejería) y al trabajo en equipo que ha hecho posible que hablen el mismo idioma organismos tan diferentes como son la Consejería de salud, el SAS y el IECA.



Equipo COVID de Andalucía

Tratamiento de los datos

En el Servicio de Estadísticas de la Consejería se normalizan los datos (formateo de variables para corregir espacios en blanco, caracteres extraños, etc.), también se realizan controles de calidad en cuanto a fechas válidas (cálculo de fechas faltantes como inicio de síntomas, diagnóstico... y agrupación de dichas fechas por semana), edades, año, sexo, territorio (recodificación de Distritos, municipios vacíos), etc., también se codifican las variables para permitir la explotación posterior de los datos.

Todo esto en principio se hacía de forma más manual, trabajando con Access y poco a poco se fue incrementando el uso de R para ir automatizando los procesos.

Incorporación de los datos a Badea

En Badea la información se encuentra organizada por cubos, cada cubo es un conjunto de datos clasificados en base a unas variables concretas.

En principio se diseñaron dos cubos, uno que recogía la serie histórica de datos agrupados por provincias y otro más detallado con información a nivel municipal que se actualizaba cada día, ambos todavía con pocos indicadores. Conforme cambiaban los criterios en el Ministerio, se iban cambiando los indicadores y ampliándose en número, así aparecen por

ejemplo los confirmados PDIA (confirmados en la Prueba Diagnóstica de Infección Activa).

En cuanto a la dimensión territorial, los criterios también fueron cambiando. Se empezó ofreciendo datos provinciales (según mapa sanitario), luego datos por distrito sanitario y finalmente por municipio.

Posteriormente el cubo de información municipal fue creciendo y en la actualidad tiene 25 dimensiones: sexo, edad, fecha de diagnóstico, fecha de declaración, fecha de defunción, etc., y cerca de 30 medidas: datos totales, a 7 días, a 14, etc.

Se avanzó con la explotación de los casos por los distintos tipos de fechas, que se recogían en un nuevo cubo, el cual permitió mostrar gráficos de evolución.

En junio de 2020 se organizó la información para poder facilitar ya no las medidas concretas, sino las medidas en base a las cifras de población de cada territorio, de manera que se facilitaron las tasas de confirmados PDIA y total confirmados.

En noviembre de 2020 se organiza la información para mostrar la evolución de las tasas de incidencia por provincia y por sexo.

En un principio, los datos se hacían públicos cada día, incluyendo fines de semana y festivos. La publicación se hacía por la tarde, una vez que el Ministerio había hecho su publicación, lo que hizo necesario pasar de una a siete las personas que en el Servicio tenían acceso a Badea para poder ir rotando.

En junio de 2020 el Ministerio acuerda publicar solo de lunes a viernes sobre las 13:00h y el lunes incluye los datos del sábado y domingo.

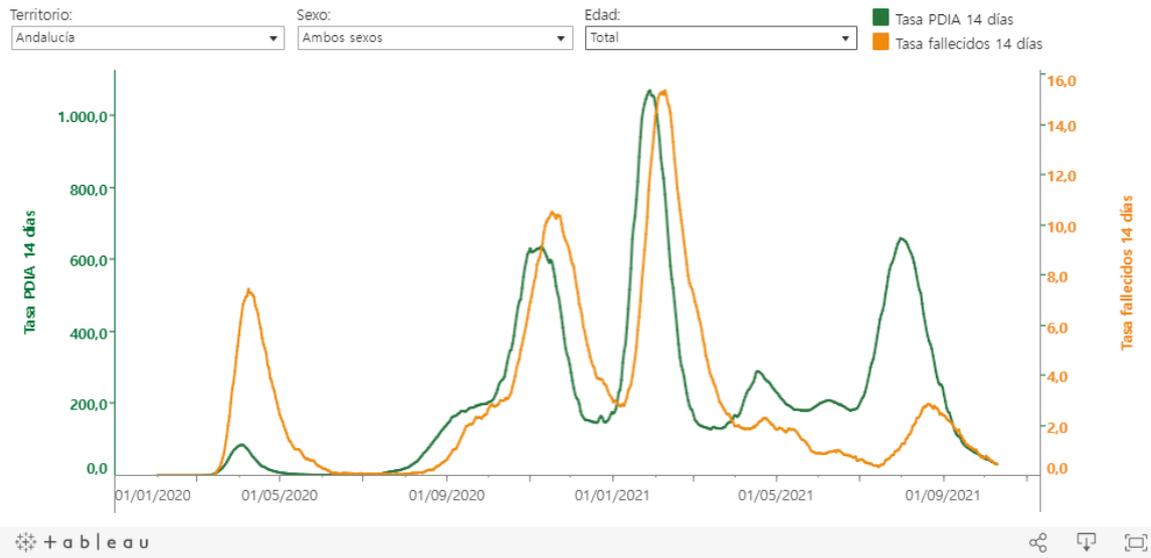
Poco después, se empieza a publicar de forma independiente a la publicación de datos por parte del Ministerio.

Desde abril de 2021 dejan de ofrecerse los datos de los domingos, por lo que el lunes se carga solamente información del sábado y del lunes.

Como consecuencia de la publicación de los datos de vacunación en abril de 2021, se ha incorporado en muchas consultas el sexo y la edad según los tramos de edad establecidos en la estrategia de vacunación.

En agosto de 2021 se crea un nuevo cubo que permite calcular en cada fecha la evolución a 14 días de las tasas de incidencia y de fallecidos. Mediante Tableau se pueden representar gráficos con doble eje que reflejan el cambio de tendencia como consecuencia de la vacunación (y que sirve como base a la estimación de muertes evitadas gracias a las vacunas).

Evolución de las tasas de incidencia PDIA y de fallecidos en los últimos 14 días a 11/10/2021



Elaborado por: Instituto de Estadística y Cartografía de Andalucía
Fuente: Dirección General de Salud Pública y Ordenación Farmacéutica. Sistema de Vigilancia Epidemiológica de Andalucía. Consejería de Salud y Familias

4. Resultados

Como resultado más palpable de todo este trabajo tenemos el ‘Portal IECA sobre el COVID-19 en Andalucía’, que fue diseñado por la Consejería de Salud y el Servicio de Difusión del IECA.

Dicho portal consta de varias secciones una de ellas, ‘Datos sanitarios’, es la que recoge toda la información relacionada con la evolución y el comportamiento de la pandemia. Inicialmente contenía la publicación ‘Datos Coronavirus (COVID-19) en Andalucía’, que es un árbol organizado de consultas y gráficos y el ‘Informe COVID-19 en Andalucía’, que a modo cuadro de mando sintetiza la situación epidemiológica a través de un conjunto de tablas, gráficos y mapas, que permiten al usuario interactuar con la información.

Más adelante, en abril de 2021, se completó esta sección con ‘Datos Vacunación COVID-19 en Andalucía’ y con el ‘Informe de Vacunación Covid-19 en Andalucía’.

El resto de secciones son un compendio de consultas de temas relacionados con la pandemia, como información de coyuntura económica, empresas, información sociodemográfica, etc.

Y a punto estuvimos de morir de éxito...

Un hito importante en la publicación de los datos del COVID 19 tuvo lugar el 27 de abril de 2020, fecha en la que empezamos a ofrecer información a escala municipal. Al día siguiente, empezamos a tener problemas de memoria.

Ha sido muy motivador ver como nuestro trabajo ha sido la base para muchas decisiones que se han tomado en el ámbito político y sanitario.

Satisfacción también al ver que todos los días estábamos presentes en los diferentes medios de comunicación, en redes sociales y en boca de la gente. Más de 8 millones de andaluces estaban pendientes de las cifras que “escupía” nuestro sistema de información, BADEA.

También destacar como algo positivo que otras Consejerías han visto como la Consejería de Salud y Familias, gracias a BADEA, ha sido capaz de ofrecer de forma rápida, organizada y transparente toda la información de la pandemia y éste debería ser un acicate para que imiten lo hecho y carguen y difundan toda su información a través de BADEA.

Y dado que el IECA se ha colado en las casas de los andaluces, lamentablemente por culpa de la pandemia, tendríamos que aprovechar para dar a conocer la gran variedad de productos y servicios que ponemos a disposición de la ciudadanía, investigadores, empresas, etc., y que son de gran utilidad para multitud de cuestiones.

Otra cuestión, relacionada con el mundo de la interoperabilidad, tiene que ver con la necesidad de que la propia Administración sea capaz de integrarla en sus procesos, cosa que no siempre ocurre y hemos podido comprobar en nuestra casa, donde hemos tenido que lamentar (aún lo hacemos) la existencia de un portal publicado por la Consejería de Presidencia, Mapa-Covid, que muy a pesar nuestro no consume los servicios publicados por BADEA, lo que ha provocado en ciertas ocasiones la existencia de cifras oficiales distintas accesibles a la ciudadanía.

Ya para finalizar, decir que esta experiencia nos ha servido para aprender y hacer un diagnóstico de nuestras debilidades con miras a preparar toda nuestra infraestructura física y lógica para convertirnos en el futuro en la OFICINA DEL DATO.

6. Principales referencias bibliográficas

Portal IECA sobre el COVID-19 en Andalucía

<http://www.juntadeandalucia.es/institutodeestadisticaycartografia/salud/index.htm>