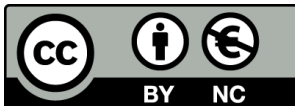


STATISTICAL METHODS TO MANAGE AND
ANALYSE REAL WORLD DATA. DEVELOPMENT
OF A HEALTH OBSERVATORY IN ALT EMPORDÀ

MÈTODES ESTADÍSTICS PER GESTIONAR I
ANALITZAR REAL WORLD DATA.
DESENVOLUPAMENT D'UN OBSERVATORI DE
SALUT A L'ALT EMPORDÀ

Álvaro Franquet Bonet

Per citar o enllaçar aquest document:
Para citar o enlazar este documento:
Use this url to cite or link to this publication:
<http://hdl.handle.net/10803/673830>



<http://creativecommons.org/licenses/by-nc/4.0/deed.ca>

Aquesta obra està subjecta a una llicència Creative Commons Reconeixement-NoComercial

Esta obra está bajo una licencia Creative Commons Reconocimiento-NoComercial

This work is licensed under a Creative Commons Attribution-NonCommercial licence



DOCTORAL THESIS

**Statistical methods to manage and
analyse *Real World Data*.**

**Development of a health observatory
in Alt Empordà**

Álvaro Franquet Bonet

2021



DOCTORAL THESIS

**Statistical methods to manage and
analyse *Real World Data*.**

**Development of a health observatory
in Alt Empordà**

This thesis contains appendices with the codes used and the infographics obtained as a result of it. The thesis is presented in English and Catalan with the intention of facilitating its dissemination.

Álvaro Franquet Bonet

2021

Doctoral Programme in Molecular Biology, Biomedicine and Health

Under the supervision of:

Dr. Maria Antònia Barceló Rado and Dr. Pere Plaja Roman

Thesis delivered to obtain the title of doctor from the University of Girona



La **Prof. Dr. Maria Antònia Barceló Rado**, de la Universitat de Girona i membre del Grup de Recerca en Estadística, Econometria i Salut (GRECS) i el **Dr. Pere Plaja Roman**, de la Universitat de Girona i director d'Innovació i Recerca de la Fundació Salut Empordà

CERTIFIQUEN:

Que el treball titulat "Statistical methods to manage and analyse Real World Data. Development of a health observatory in Alt Empordà", que presenta Alvaro Franquet Bonet per a l'obtenció del títol de doctor, ha estat realitzat sota la seva direcció i que compleix els requisits per poder optar a Menció Internacional i a Menció Industrial.

I, perquè així consti i tingui els efectes oportuns, signem aquest document.

Girona, 2 de juliol de 2021

ACKNOWLEDGEMENTS

Thank you to the supervisors of this thesis. To **Pere** for listening and for trusting in the proposals that emerged during the process. To **Maria Antònia** for her patience and commitment.

Thank you to the people at **Fundació Salut Empordà**, and especially **Eneida** and **Lluís** for all those moments in the office, for all the fun we had over our mid-week lunches and for all the work they put into ensuring that this thesis received funding. Thank you to **Berta** for the corrections and for making sure everything went well. Thank you to **Jordi, Sergi, Ferran i Forment** for the laughs we had in the office.

Thanks to the people in **GRECS** who made me feel at home, even though they were not often in the office, and especially to **Marc** for his commitment and to **Marta** always answering my questions with a smile.

Thanks to **Rut** and **Ivonne** for the wonderful **Erasmus University Rotterdam** experience, where I had the pleasure of helping them write the paper “HOW TO CREATE THE MOST COMPLETE LISTING OF OVERALL HAPPINESS IN NATIONS: Methods for estimating average life-satisfaction in nations/years where questions on that matter have not been asked” with you.

Thanks to the people at **Universitat Politècnica de Valencia** for accepting me onto the Bayesian statistics course and for the experience of being with them for its duration.

*Thanks to the people at Escuela Andaluza de Salud for looking after us so well during the meetings and the conference. To **Andrés** for sharing his experience in creating health observatories.*

Last, thanks to all the marvellous people, both **friends and family**, who have helped me in many and diverse ways to follow this path and to achieve what I set out to do. Thanks to **Albert** to **Katerina** for putting me up and making me feel at home. Thanks to **Carmesina**, simply for *being* at all times and in all possible

ways. Without your patience, this thesis would not have been possible. And thanks to my mum Ana, my brother Marc and aunt Pili for being there for me so much and so well these 30 years. And last, thanks to those who are no longer here but who would certainly have been as happy for me as the ones who are here.

Thank you very much eve!

Álvaro

FUNDING

This doctoral thesis received financial support from the 'Industrial Doctorates (DI 2018)' call. First partial definitive resolution of the industrial doctorates call (DI) 2018 of the president of the Executive Commission for University Research Grants, 10 September 2018. Thesis supervisors: Maria Antònia Barceló (GRECS, UdG) and Pere Plaja (Fundació Salut Empordà).

Funding was also received from the European Regional Development (FEDER) associated with the public subsidies call of the European Services Programme, published in the Official Bulletin of Girona no. 94 of 18 May 2016.

It is associated with the development part of the Health and Social Pole of the Alt Empordà, promoted by the Fundació Salut Empordà within the framework of the "Girona, healthy region" project (European Territorial Collaboration Project, PECT), in which Girona Provincial Council, Dipsalut and IdibGi also took part.

LIST OF ABBREVIATIONS

API *Application Programming Interface*

AQuAS *Agency for Health Quality and Assessment of Catalonia*

CDM *Common Data Model*

COVID-19 *Coronavirus Disease 2019*

CRAN *Comprehensive R Archive Network*

CSV *Comma-separated value*

cURL *Client Uniform Resource Locator*

DGT *Traffic Agency*

Dipsalut *Public Health Body of Girona Provincial Council*

DM2 *Diabetes Mellitus type II*

ECAP *Primary Healthcare Information System*

EDA *Event Driven Architecture*

ELISA *Enzyme-linked immunoassay*

EUROSTAT *European Statistics Office*

FSE *Empordà Health Foundation*

GPL *General Public License*

GRECS *Research Group in Statistics, Econometrics and Health*

HTML *HyperText Markup Language*

List of abbreviations

IAEG-SDG *Inter-agency and Expert Group on Sustainable Development Goals Indicators*

Idescat *Statistical Institute of Catalonia*

IdiBGI *Dr Josep Trueta Institute of Biomedical Research*

INE *National Statistics Institute*

JSON *JavaScript Object Notation*

OBSA *Asturias Health Observatory*

ODS *Sustainable development goals*

OHDSI *Observational Health Data Sciences and Informatics*

OKF *Open Knowledge Foundation*

OMS *World Health Organisation*

ONU *United Nations Organisation*

OpenEHR *Open Electronic Health Records*

PADRIS *Public data analysis programme for health research and innovation*

PCR *Polymerase chain reaction*

PECT *Territorial Competitiveness Specialisation Project*

RCA *Central Registry of Insured Persons*

RCT *Randomized clinical trial*

RWD *Real World Data*

RWE *Real World Evidence*

SDH *Social Determinants of Health*

List of abbreviations

UCI *Intensive Care Unit*

UdG *University of Girona*

URL *Uniform Resource Locator*

W3C *World Wide Web Consortium*

WHO *World Health Organization*

WU *Wirtschaftsuniversität Wien*

XML *Extensible Markup Language*

LIST OF FIGURES

Figure 1 Dahlgren i Whitehead's model of the determinants of health. Source: Dahlgren and Whitehead (1991).	3
Figure 2 Conceptual framework of the Social Determinants of Health. Source: Comisión para reducir las desigualdades en salud en España (2012).	4
Figure 3 Working model proposed by OBSA. Source: Author's own elaboration.	11
Figure 4 Model of the process of creating infographic. Source: Author's own elaboration.	20
Figure 5 Sketch of an infographic. Source: Author's own elaboration.	22
Figure 6 JSON object schema. Source: Author's own elaboration based on https://www.json.org/json-en.html	29
Figure 7 Definition schema of a matrix using JSON. Source: Author's own elaboration based on https://www.json.org/json-en.html	29
Figure 8 Schema of a label in XML format. Source: Author's own elaboration.	30
Figure 9 Schema of an element in XML format. Source: Author's own elaboration.	31
Figure 10 Schema of an element in XML format with attributes. Source: Author's own elaboration.	31
Figure 11: Flux diagram of an event driven architecture. Source: Author's own elaboration.	41
Figure 12 Three drafts for the "Population in Alt Empordà" Infographic. Author's own elaboration.	48
Figure 13 " Població a l'Alt Empordà " Infographic version 2019. Author's own elaboration.	49
Figure 14 Draft of the infographic "Birth rate in Alt Empordà" and the final infographic. Author's own elaboration.	50

Índex de figures

Figure 15 Draft of the infographic "Un any de COVID-19 a l'Alt Empordà". Source: Author's own elaboration.....	51
Figure 16 Prototype of the infographic "Un any de COVID-19 a l'Alt Empordà". Source: Author's own elaboration.....	52
Figure 17 Infographic "Un any de COVID-19 a l'Alt Empordà". Source: Author's own elaboration.....	52
Figure 18 Screenshot of the daily Covid-19 data by county of the Health Department. Author's own elaboration.	53
Figure 19 Application "Situation of COVID-19 in Alt Empordà". Source: Author's own elaboration.....	58
Figure 20 Front cover of the Key Indicators of Alt Empordà 2020. Source: Author's own elaboration	60
Figure 21 Downloads of the eventr package from 2020-05-26 to 2021-06-04. Source: Author's own elaboration based on data obtained through the cranlogs package.....	85
Figure 22 Home page of the Indika data repository. Source: Author's own elaboration.....	86

LIST OF TABLES

Table 1 List of observatories analysed in the definition of Indika.....	9
Table 2 Attention attributes prior to visual perception.....	23
Table 3 Licenses that are suitable for open data.....	26
Table 4 Classification of open data using a system of stars.....	27
Table 5 Meta information of the indicators on the repository.....	39
Table 6 Example of occurrences of three events. Source: Author's own elaboration.....	42
Table 7 Daily Covid-19 data by county.....	54
Table 8 Structure of the Table "Registry of COVID-19 cases in Catalonia"	57
Table 9 Traffic on the Indika website between January and June.....	61
Table 10 Number of monthly downloads of the eventr package from the CRAN repository. Source: Author's own elaboration based on data obtained through the cranlogs package.	85

ABSTRACT

The objective of this thesis is to explore in greater depth the process of creating a public health observatory and the processing of the information needed to take territorial decisions in the county of Alt Empordà. The outcome is the creation of Indika, the Health and Social Pole of the Alt Empordà. Indika is a public health agency that acts as the territory's health and well-being agent. The objective of Indika is to generate knowledge about the impact of social determinants of health in Alt Empordà. To do so, it uses a working framework based on three points: to inform, to discuss, and to act. This working framework served to produce the visual information solutions, including infographics, collections of indicators, and a web application whose purpose is to portray the existing problems in the territory to start debate with the different political and social agents to take decisions on improvement initiatives.

Apart from creating a public health observatory, two parallel lines of action have also been developed in this thesis: the creation of public structures to store information based on events as a way of structuring medical information. These two lines of investigation have resulted in the creation of the "Indika Data Repository", an information repository about the county of Alt Empordà, using the collaborative platform GitHub and the implementation of the eventr package in R programming language, and currently published in the Comprehensive R Archive Network. Eventr is a package whose purpose is to facilitate the implementation of architectures based on events.

Key words: public health observatory, data repository, Real World Data, architecture based on events.

RESUM

Aquesta tesi té com a objectiu aprofundir en el procés de creació d'un observatori de salut pública i el tractament de la informació necessària per a la presa de decisions territorials a la comarca de l'Alt Empordà.

El resultat ha estat la creació d'Indika pol de salut i social de l'Alt Empordà. Indika és un observatori de salut pública que actua com a agent de salut i benestar del territori. L'objectiu d'Indika és generar coneixement sobre l'impacte dels determinants socials de salut a l'Alt Empordà. Per a això fa ús d'un marc de treball basat en tres punts: informar, conversar i actuar.

Aquest marc de treball s'ha concretat en la generació de visualitzacions d'informació com infografies, col·leccions d'indicadors i una aplicació web amb l'objectiu de donar a conèixer les problemàtiques del territori i iniciar així un debat amb els diferents actors polítics i socials que permeti concretar accions de millora.

En aquesta tesi, a més de la creació d'un observatori de salut pública, s'han desenvolupat dues línies d'investigació paral·leles que són: la creació d'estructures públiques d'emmagatzematge d'informació i l'anàlisi de l'arquitectura basada en esdeveniments com a forma d'estructurar la informació mèdica.

Aquestes dues línies d'investigació s'han concretat en la creació de "Indika Data Repository", un repositori d'informació sobre la comarca de l'Alt Empordà a través de la plataforma col·laborativa GitHub i la implementació de la llibreria eventr en el llenguatge de programació R publicada actualment al Comprehensive R Archive Network. Eventr és una llibreria que té com a objectiu facilitar la implementació d'arquitectures basades en esdeveniments.

Paraules clau: observatori de salut pública, repositori de dades, Real World Data, arquitectura basada en esdeveniments.

RESUMEN

Esta tesis tiene como objetivo profundizar en el proceso de creación de un observatorio de salud pública y el tratamiento de la información necesaria para la toma de decisiones territoriales en la comarca del Alt Empordà.

El resultado ha sido la creación de Indika Polo de salud i social de l'Alt Empordà. Indika es un observatorio de salud pública que actúa como agente de salud y bienestar del territorio. El objetivo de Indika es generar conocimiento sobre el impacto de los determinantes sociales de salud en el Alt Empordà. Para ello hace uso de un marco de trabajo basado en tres puntos: informar, conversar y actuar.

Este marco de trabajo se ha concretado en la generación de visualizaciones de información como infografías, colecciones de indicadores y una aplicación web con el objetivo de dar a conocer las problemáticas del territorio e iniciar así un debate con los diferentes actores políticos y sociales que permita concretar acciones de mejora.

En esta tesis, además de la creación de un observatorio de salud pública, se han desarrollado dos líneas de investigación paralelas que son: la creación de estructuras públicas de almacenamiento de información y el análisis de la arquitectura basada en eventos como forma de estructurar la información médica.

Estas dos líneas de investigación se han concretado en la creación de "Indika Data Repository", un repositorio de información sobre la comarca del Alt Empordà a través de la plataforma colaborativa GitHub y la implementación de la librería eventr en el lenguaje de programación R publicada actualmente en el Comprehensive R Archive Network. Eventr es una librería que tiene como objetivo facilitar la implementación de arquitecturas basadas en eventos.

Palabras clave: observatorio de salud pública, repositorio de datos, Real World Data, arquitectura basada en eventos.

INDEX

ACKNOWLEDGEMENTS.....	III
FUNDING	VII
LIST OF ABBREVIATIONS.....	IX
LIST OF FIGURES.....	XIII
LIST OF TABLES	XV
ABSTRACT	XVII
RESUM	XIX
RESUMEN	XXI
INDEX	XXIII
1 Background.....	1
1.1 Social determinants of health	2
1.2 Sustainable development goals	6
1.3 Public health observatory	8
1.4 Real world data	12
1.5 Event driven data models.....	17
1.6 Visualisation of information	19
1.6.1 Characteristics of a visualisation.....	19
1.6.2 Methodology in the process of creating an infographic.....	20
1.6.2.1 Strategy.....	20
1.6.2.2 Data	21
1.6.2.3 Design.....	21
1.6.3 Principles of an information infographic.....	22
1.7 Open Data.....	25
1.7.1 Open data classification system	27

Index

1.7.2	JavaScript Object Notation	28
1.7.3	Extensible Markup Language (XML).....	30
2	HYPOTHESIS AND OBJECTIVES	33
2.1	Hypothesis	33
2.2	Objectives	33
3	Methods	35
3.1	Design a repository of health, socioeconomic and environmental indicators	35
3.1.1	Identification of requirements.....	35
3.1.1.1	Review of indicators	36
3.1.1.2	Technical requirements.....	37
3.1.2	Design and architecture of the repository	38
3.2	Construction of a health, socioeconomic and environmental repository for Alt Empordà.....	38
3.2.1	API.....	40
3.3	Use of the integrated data bases	40
3.3.1	Key Indicators of Alt Empordà	40
3.4	Implementation of an R package for modelling data based on events	40
3.4.1	Terminology.....	41
3.4.2	Implementation	42
4	Results	47
4.1	Indika Health and Social Pole of the Alt Empordà.....	47
4.1.1	Infographics	47
4.1.1.1	Population in Alt Empordà.....	47
4.1.1.2	The “La Natalitat a l’Alt Empordà ” Infographic	49
4.1.1.3	“Un any de COVID-19 a l’Alt Empordà” Infographic.....	50
4.1.2	Application on the COVID-19 situation in Alt Empordà	52

Index

4.1.3	Key Indicators for Alt Empordà	59
4.1.4	Website of the Indika project	61
4.1.5	Other media	61
4.2	eventr of R package	61
4.2.1	Structure of the individual	62
4.2.2	Create the specifications.	62
4.2.3	Creating the handlers	64
4.2.4	Create the dispatcher,	66
4.2.5	Creating the occurrences of the events	67
4.2.6	Executing the events	68
4.2.7	Advantages of the use of EDA.....	69
4.2.7.1	Obtaining the auditory registry	69
4.2.7.2	Testing the events	70
4.2.7.3	Retroactive events	72
4.2.7.4	Alternative states.....	73
4.2.8	Application to a real case.....	74
4.2.8.1	Structure of the individual.....	75
4.2.8.2	Create the specifications of the events.	76
4.2.8.3	Create the handlers.....	77
4.2.8.4	Creating the dispatcher	79
4.2.8.5	Creating the occurrences of the individuals.....	79
4.2.8.6	Execute the events for each individual.....	81
4.2.8.7	Obtain the Audit Trail	83
4.3	Indika data repository	86
5	Discussion	89
5.1	Indika Health and Social Pole of the Alt Empordà	89

Index

5.2	Indika data repository	92
5.3	Eventr.....	92
6	Conclusions and future work.....	95
7	References.....	97
	Appendices.....	108
	<i>Code of the “Population in Alt Empordà” infographic</i>	108
	Població a l’Alt Empordà infographic	110
	<i>Code of the “La Natalitat a l’Alt Empordà” infographic</i>	111
	La Natalitat a l’Alt Empordà infographic.....	115
	<i>Code of the “Un any de COVID-19 a l’Alt Empordà” infographic</i>	116
	Un any de COVID-19 a l’Alt Empordà infographic	120
	<i>Code of the “Un any de COVID-19 a l’Alt Empordà” infographic</i>	122
	<i>Key Figures on Europe</i>	126
	list of key indicators for Alt Empordà.....	137
	Press releases	144

1 BACKGROUND

The background of this thesis are the European territorial collaboration project (PECT) "Girona, regió saludable" [Girona, healthy region"], implemented by the Empordà Health Foundation (FSE), Girona provincial council through the Public Health Body of the Province of Girona (Dipsalut) and the Biomedical Research Institute of Girona (IdIBGi), and an industrial doctoral project in which the Research group in Statistics, Econometrics and Health (GRECS) of the University of Girona (UdG) and the FSE collaborate, culminating in the presentation of this thesis.

The FSE is a non-profit foundation, the mission of which is to provide comprehensive services in the area of health and related issues, geared towards the needs of the population in the Girona county Alt Empordà. The FSE provides primary healthcare services (Escala basic health area), specialised care (Figueres Hospital), and socio-health care (Bernat Jaume Socio-health Centre).

In the last years, and driven by the will of different organisations, one of which is the FSE, various initiatives have been launched aimed at accumulating more knowledge about the reality of the county of Alt Empordà in relation to state of health and the social determinants that condition it (Dahlgren and Whitehead, 1991), taking the Sustainable Development Goals as the guiding principle (ODS) (Griggs, et al., 2013).

The county of Alt Empordà is made up of 68 municipalities and a total population of 141,339 people (Idescat, 2019a), 46,654 (33%) of which are registered in municipal registry of Figueres (Idescat, 2019b), the capital of the county and the headquarters of the foundation, the hospital, and the socio-health centre. The organisation provides health care for a total of 141,539 inhabitants. The difference between the number of inhabitants and the people that receive healthcare is due to the inclusion of the population living in the municipalities of Albons and Bellcaire, which are not in fact in Alt Empordà.

Background

The FSE would like to place special emphasis on the use of the data and decision making based on scientific evidence to improve the health of the population. Based on this vision, and with the aim of improving the collection of information and the transferal of this knowledge to the politicians to make critical decisions in the country, the UdG and GRECS of the UdG have together promoted the Health and Social Pole of the Alt Empordà, a health and social observatory whose strategy stems from the territory, and a working model based on three fundamental lines of action inform, discuss and act (Cofiño *et al.*, 2012).

The main aim of Indika, the Health and Social Pole of the Alt Empordà is the ongoing improvement of the health and quality of life of the population, paying special attention to reducing social health inequalities (World Health Organization, WHO, 2010). To do so, the social determinants of health (SDH) are used as tools to measure progress in the SDG (United Nations General Assembly, ONU, 2015).

1.1 SOCIAL DETERMINANTS OF HEALTH

SDH are the non-medical factors that influence state of health. They are the conditions in which one is born, grows, works, and lives, plus age and other factors that have an impact on people's daily lives. These factors include economic policies, social norms, and political systems, among others (WHO, 2020). Various studies have shown how the social determinants of health are relevant to increased rates of death and illness (Wilkinson i Marmot, 2003). SDH include a very wide set of factors such as life conditions, political systems, and economic, cultural and power structures (WHO, 2020).

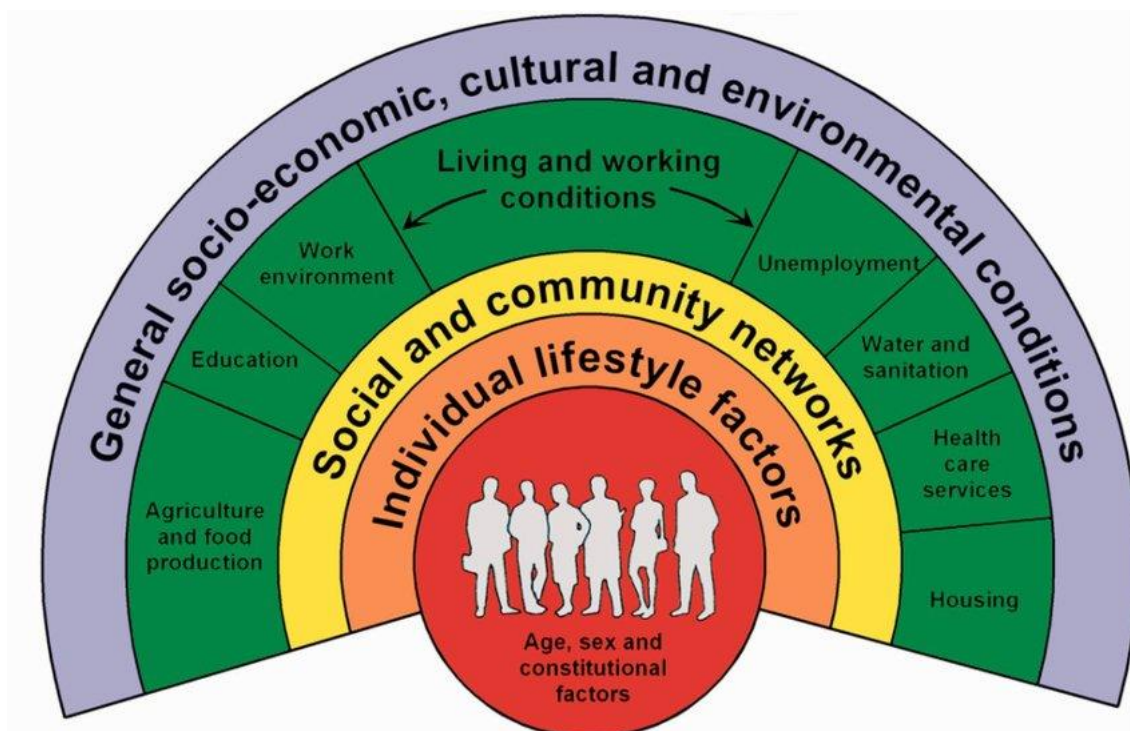


Figure 1 Dahlgren i Whitehead's model of the determinants of health. Source: Dahlgren and Whitehead (1991).

Dahlgren and Whitehead (1991) propose a model of the determinants of health structured as a rainbow made up of multiple layers of influence (Figure 1). At the centre of the figure, individuals have an age, a gender, and constitutional characteristics that are largely fixed. Around them are the influences that are theoretically modifiable by policies. First, there are personal behavioural factors such as smoking and physical activity, while second, the individuals interact with their friends and the immediate community and are influenced by them, which is represented in the second layer. The next layer is a person's capacity to stay healthy, which is influenced by their life and work conditions, their food supply, and their access to essential goods and services. Last, economic, cultural, and environmental influences in society in general act as mediators of the health of the population. This model to describe the determinants of health places the emphasis on interaction: individuals' lifestyles are embedded in social norms and networks and in life and working conditions, which in turn are related to the wider socio-economic and cultural world (Dahlgren and Whitehead, 2006).

Background

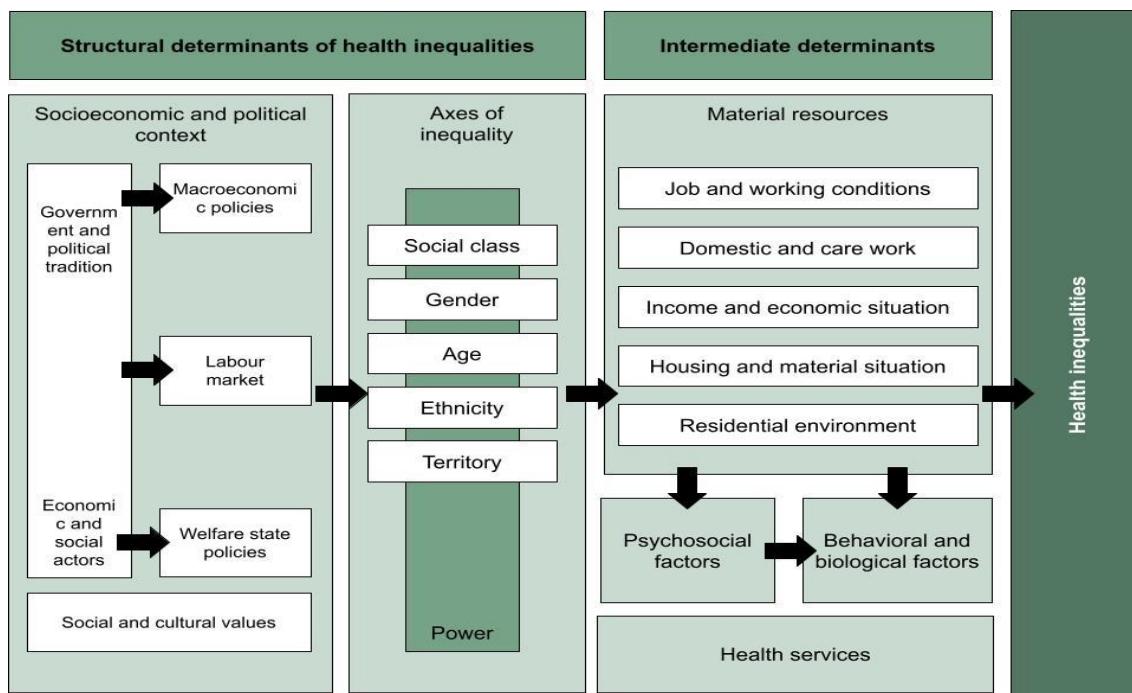


Figure 2 Conceptual framework of the Social Determinants of Health. Source: Comisión para reducir las desigualdades en salud en España (2012).

The “Comisión para reducir las desigualdades en salud en España” (2012) proposes another model or conceptual framework upon which to base decision making (Figure 2). This conceptual framework consists of two main elements: the structural elements and the intermediate factors of health inequalities. The environmental factors are the socioeconomic and political context, and the axes of inequality. The socioeconomic and political context includes the government and other economic factors that impact general state of well-being, and economic and labour market policies, among others. Also included in this sub-section are the social and cultural values that provide the philosophical and moral basis upon which to underpin political decisions and the society’s power hierarchies. The axes of inequality consider the different social structures that can affect individual opportunities to have good health. Some of these social structures are the territory, ethnicity, age, gender, and social class.

The intermediate determinants include material resources, psychosocial factors, behavioural and biological factors, and health services. Material resources can be subdivided into the following five groups: (1) job and working conditions, (2) domestic and care work, (3) income and economic situation, (4) housing and material situation, and (5) residential environment (Daponte *et al.* 2019).

Background

Social inequalities are systematic health differences among the different socioeconomic groups of a country. The differences often negatively affect the least socially privileged groups for socioeconomic reasons, age, or ethnicity. Given that there is no biological reason why they exist, it is evident that they are largely shaped by structural, socioeconomic, and cultural factors, for which reason they must be addressed (Whitehead, 2006).

All the systematic differences in health between a country's various socioeconomic groups can be considered unjust and, therefore, can be classified as health inequalities. There is no biological reason for their occurrence, and it is evident that even the systematic differences in lifestyles between economic groups are largely shaped by structural factors. In summary, the social inequalities in health are directly or indirectly generated by social, economic, or environmental factors, and structurally affected lifestyles. These determinants of social inequalities can change.

The commission of the social determinants of health was created in 2005 by the WHO and global health promotion associations to address the social factors that lead to ill health and inequalities. The commission identified social determinants of health as one of the main causes of ill health and inequalities in countries and between them.

In 2008, the commission published the book "*Closing the gap in a generation: health equity through action on the social determinants of health - Final report of the commission on social determinants of health*" (WHO, 2008). In the book, the commission recommends focusing action on three key points, (1) improving living conditions, (2) fighting against the unequal distribution of power, money, and resources (3) measuring the magnitude of the problem with the aim of analysing and evaluating the effects of the interventions carried out.

The inequalities caused by the way in which society is organised mean that the possibilities to develop one's life and enjoy good health are badly distributed both within a society and between the different societies. These inequalities can be seen in the rate of schooling in infancy, types of jobs and working conditions, and the characteristics of the constructed environment, and the quality of the

Background

environment where the population lives. Depending on the characteristics of this environment, physical conditions, psychosocial support, and behavioural norms vary for each group, making them vulnerable to illness to varying degrees. Social stratification also creates disparities in access to the health system and its use, leading to inequalities in the promotion of health and well-being, the prevention of diseases, and the possibilities of recovery and survival after suffering a disease (WHO, 2008).

Inequality in life conditions is determined by more deeply entrenched structures and social processes. The inequalities are systematic and are the result of social norms, policies, and social practices that tolerate and even facilitate an unfair distribution of power, wealth, and other required social resources, and access to them.

Action on the social determinants of health would be more effective if there were basic data systems like the civil registers and the systematic health inequalities and social determinants of health observation programmes, and mechanisms to ensure that the data can be interpreted used to formulate policies and set up more effective systems and programmes. Awareness-raising and training on the matter of determinants of health is essential.

1.2 SUSTAINABLE DEVELOPMENT GOALS

The Sustainable Development Goals (SDG), also known as the Global Goals, were adopted by all the member states in 2015 as a universal call to end poverty, protect the planet, and guarantee that everybody enjoys peace and prosperity by 2030 (UN General Assembly, 2015). The 17 SDG are integrated, recognising that interventions in one area will affect the results of the others and that development must be balanced with environmental, economic, and social sustainability. To fulfil the pledge that nobody is left behind, the countries have promised to accelerate progress for those lagging most. To this end, the SDG have been designed to bring various “zeros” to the world which will change people’s lives and include zero poverty, zero starvation, zero AIDS, and zero discrimination against women and children. Everyone needs to help achieve these ambitious

Background

objectives. What is needed is creativity, knowledge, technology, and the financial resources of all of society to achieve the SDG in each context.

The 17 SDG are: (1) an end to poverty, (2) zero starvation, (3) good health and well-being, (4) quality education, (5) gender equality, (6) clean water and sanitation, (7) affordable and clean energy, (8) decent work and economic growth, (9) industry, innovation, and infrastructure, (10) reduced inequalities, (11) sustainable cities and communities, (12) responsible consumption and production, (13) climate action, (14) life below water, (15) life on land, (16) peace, justice and strong institutions, (17) partnerships for the goals.

In relation to the SDG, in the 48th session of the UN Statistics Commission in March 2017, the Inter-institutional and SDG Experts Group (IAEG-SDG) developed a framework of global indicators, which was later adopted by the General Assembly on 6 July 2017 and can be found in the Resolution adopted by the General Assembly on Employment of the Statistics Commission in relation to the 2030 Agenda for Sustainable Development (A / RES / 71/313), Appendix (ONU, 2020). According to the Resolution, the framework of indicators was planned to be exhaustively clarified and revised by the Statistics Commission in its 51st and 56th sessions in March 2020 in 2025, respectively. The framework of global indicators will be complemented with regional and national level indicators developed by the member states.

The yearly clarifications made to the indicators will be progressively included in the framework of indicators as they are produced. In line with the group mandate, the IAEG-SDG proposed 36 important changes to the framework in the form of substitutions, revisions, additions, and eliminations as part of the 2030 comprehensive review, which was approved by the 51st Statistics Commission in the March (ONU, 2020).

The framework of global indicators includes 231 unique indicators. However, the total number of SDG indicators that feature in the framework of global indicators is 247 because 12 indicators are repeated under two or three different objectives. (ONU, 2020).

Background

1.3 PUBLIC HEALTH OBSERVATORY

A public health observatory can be understood as a virtual entity focused on the production of regular and comprehensive observations on the population, with the aim of supporting and guiding effective actions to improve public health (Castillo, 2015).

There are different types of public health observatories depending on their mission, composition, focus or geographical cover (Jacobson and Castillo, 2014). The functions of an observatory vary according to the local context in which they have the capacity to operate, and the available resources, skills, and capabilities. Another important factor will be their state of development (Jacobson and Castillo, 2014). Despite possible differences in the definition of the observatory concept, there are some characteristics that tend to be common to all the different types, for example, they generate information relevant to the moment on the state of health of a territory and they report on population trends in relation to specific SDG. They also analyse and interpret data related to the object population, and they work collaboratively with the political entities and decision makers.

Based on the work of Caiaffa *et al.* (2014), who compiled a set of surveys for different health observatories, some of these functions can be further broken down to include: the integration of local information into decision making, the provision of a general overview of health, the promotion of a culture based on specific actions and on data to help to solve local problems, to carry out analyses and scientific investigations, and to be a business intelligence unit capable of grouping data from different sources with the aim of transforming the data into useful knowledge.

In the process of designing Indika, a search of observatories that act at both the national and international administrative levels was made using two sources of information. First, the available scientific literature in different indexed journals based on the GoogleScholar and SciElo search engines; and second, the resulting information obtained from carrying out a search using the key words “*Public Health Observatory*” in the Google and DuckDuckGo search engines. The list of observatories analysed can be found in Table 1.

Background

Table 1 List of observatories analysed in the definition of Indika.

Name of the Observatory	Administrative level	URL
Global Health Observatory	World	https://www.who.int/data/gho
Asturias Health Observatory	Autonomous Region	https://obsaludasturias.com/obsa/
Observatory of Health and Policy Impact	Municipal (Barcelona)	https://ajuntament.barcelona.cat/observatorisalut/estat-de-salut/indicadors
County Health Rankings and Roadmaps	Country (USA)	https://www.countyhealthrankings.org/
Observatory of the Health System of Catalonia	Autonomous Region (Catalonia)	http://observatorisalut.gencat.cat/ca/inici
Observatory of Health Inequalities	Autonomous Region (Catalonia)	http://observatorisalut.gencat.cat/ca/observatori-desigualtats-salut/
Observatory of Mortality	Autonomous Region (Catalonia)	http://observatorisalut.gencat.cat/ca/observatori_mort/

Background

Observatory of the Results of the Madrid Health Service	Autonomous Region (Madrid)	https://www.comunidad.madrid/servicios/salud/observatorio-resultados-servicio-madrileno-salud
Observatory of Health and the Environment of Andalucía	Autonomous Region (Andalusia)	https://www.osman.es/
Mental Health Observatory	Autonomous Region (Extremadura)	https://www.feafesextremadura.com/observatorio-de-salud-mental/
Health Observatory of Valencia	Autonomous Region (Valencia)	https://www.sp.san.gva.es/sscc/opciones4.jsp?CodPunto=3707iOpcion=SANMSOVS1iMenuSup=SANMSiNivel=2
Public Health Wales Observatory	Wales (UK)	http://www.publichealthwalesobservatory.wales.nhs.uk/home
European Observatory on Health Systems and Policies	Europe	https://www.euro.who.int/en/about-us/partners/observatory
Observatories on Gender and Health in Latin America and the Caribbean	Latin America	https://www.paho.org/es

Background

“La Caixa” Social Observatory	Not specified	https://observatoriosociallacaixa.org/
New Zealand Public Health Observatory	New Zealand	www.nzpho.govt.nz

Source: Author's own elaboration

A total of 16 national and international observatories operating at different administrative levels were analysed.

The *Asturias Health Observatory (OBSA)* propose a working model (developed together with *County Health Ranking* and the University of Wisconsin) upon which to base the actions of the observatory: inform, discuss, and act (Cofiño *et al.*, 2012).

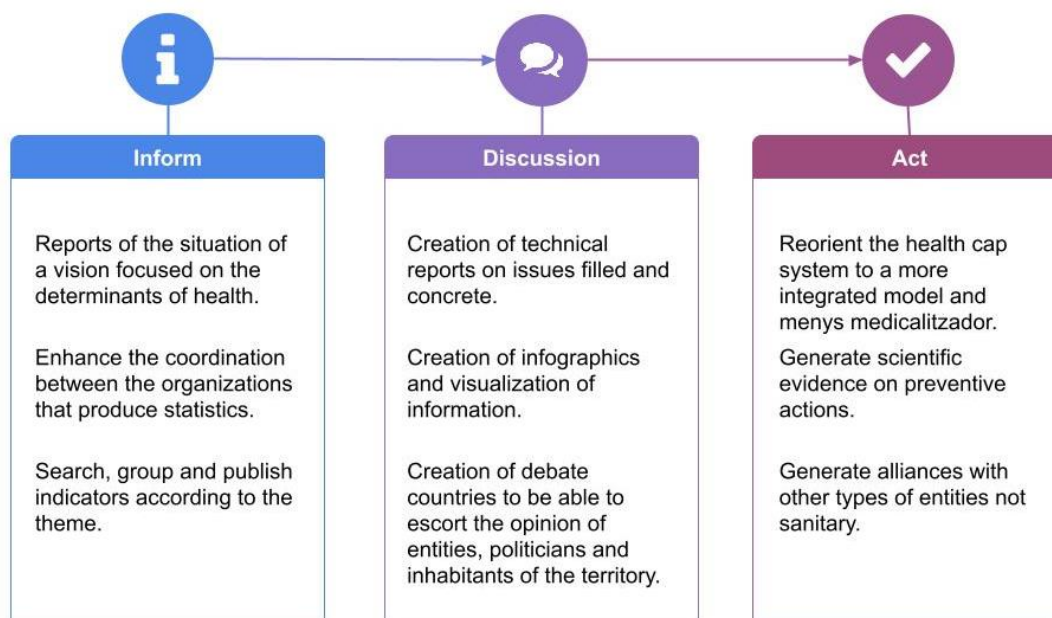


Figure 3 Working model proposed by OBSA. Source: Author's own elaboration.

The first stage (inform) consists of informing the population and the different territorial agents about the situation in the territory. The objective is to initiate discussion that facilitates decision making and helps to specify actions to improve

Background

health in the territory. To be able to inform the population accurately, it is important to know what its problems and dynamics are, and to add the perspective of its strengths and positive aspects. It is not productive to focus solely on the negative aspects of each territory; it is also important to learn its strengths.

Once the population and the agents have been informed, discussion must be initiated. There are different ways to carry out this discussion, but the most frequently used ones are the creation and diffusion of information in formats that adapt to the different sectors (like visualisations and technical reports), the creation of spaces of debate in which all the agents in the territory can take part, and training in community health matters (Cofiño *et al.*, 2012).

The last stage is to act. The objective of the actions must be to re-direct the systems, and especially the health system, towards a model that better prioritises health and is capable of generating alliances in the territory with the intention of significantly improving health in a sustainable way (Cofiño *et al.*, 2012). Given that there is still little scientific evidence regarding the different interventions, it is important to develop a scientific perspective of them to be able to generate the required evidence to ensure that they can be extrapolated to other territories and communities.

The strategy chosen by the Indika observatory was to focus on informing the population by generating reports and scientific papers jointly with the University of Girona using information visualisations like those proposed by EUROSTAT (2019, 2020) and Hametner *et al.* (2019).

1.4 REAL WORLD DATA

Real World Data (RWD) are observational studies based on real data obtained in daily clinical practice (Llano *et al.*, 2014) and from patients' daily lives. More specifically, RWD can be defined as the studies that collect data relevant to human health that do not come from random clinical trials (RCT) and conventional clinical trials (The Association of British Pharmaceutical Industry, 2011a; 2011b).

Background

Therefore, RWD combine different types of data sources: data obtained from clinical practice based on population cohorts (designed with individual data), data obtained in clinical practice based on administrative data (design with individual data), data from patients' daily lives obtained from health surveys (design with individual data), and contextual indicators about certain social and environmental health determinants (ecologic design). Real World Evidence (RWE) is understood as the evidence obtained from these studies.

RWD provide the external validity that RCT do not, whose designs have greater internal validity. In fact, the main advantage of RWD, apart from their being cheaper, the fact that they use larger samples, and providing greater retrospectivity than other research designs, is their high external validity (Staa *et al.*, 2012; Kneale *et al.*, 2016).

To achieve this internal and external validity, the evidence generated must be repeatable (meaning that the researchers must wait to produce identical results on applying the same analysis to the same data for any question), reproducible (a different researcher should be able to carry out the same task when making an analysis in a specific database and expect to produce a result that is identical to the first researcher's), replicable (the same question addressed using the identical analysis against similar data produces similar results), generalizable (circumstances in which identical analyses are made in different databases and consistently similar results are obtained), robust (meaning that the results do not have to be very sensitive to the subjective analysis that can be carried out within an analysis), and calibrated (an observational study always has to be accompanied by diagnoses that test the assumptions about the design, the methods, and the data).

There are a series of problems that impede the full leverage of the potential of RWD, among which are the lack of a standardised procedure for collecting data, the lack of quality standards of the data and their validation procedures, and the lack of representative databases and studies that show how RWD can be used in the health systems depending on their aim (for example, to take clinical and regulatory decisions, to evaluate health technologies, etc.). However, if these challenges are successfully addressed, RWD can provide new knowledge about

Background

patterns of diseases and can help to improve the security and effectiveness of medical interventions.

Obviously, RWD correspond to the investigative activity of any applied statistic. To this effect, RWD combine data that come from many sources, are standardised in a common data model, and allow standard systematic analyses to be devised.

Even though RWD have always been enormously important, it has been the progress in electronics and information technology in recent years that have allowed not only huge amounts of data to be collected but also to be processed. Over the last decade, new health information systems have been introduced on a huge scale, producing a notable increase in both the quality and the quantity of the computerised clinical data. The digitalisation and introduction of information (clinical records, patient registers, medical insurance data, health surveys, administrative data, biobanks, mobile health, etc.), supported by artificial intelligence (IA), has facilitated obtaining data in a real context (RWD) which, when analysed effectively, can generate scientific evidence (RWE).

The huge sophistication and volume of the data we currently have available has made it necessary to work with multi-disciplinary teams to be able to leverage maximum benefit and accumulate the most knowledge possible from them. Therefore, and in contrast to previously when participation was scarce, health economists, statistics, mathematics, biomedics, accountants, politicians, engineers, patients, observers, and carers, among others, have come to play important roles in terms of decision making.

Given that the data come from different sources, combining them is a hugely challenging process that includes unifying the formats, codifying the contents, and standardising the logic and the terminology used by the different organisations to describe the variables. This is further complicated by the fact that there are external data sources (data generated by mobile applications, *wearables*, data coming from of populational health surveys, contextual variables from population censuses, etc.) among the sources of RWD, and when the RWD

Background

used come from centres in different autonomous regions, countries, etc., each has its own information system.

A common data model is needed (CDM) that allows the systematic analysis of databases of disparate observational studies. There are currently different initiatives related to this concept such as, for example, OpenEHR (Kalra *et al.*, 2005) or OHDSI (Hripcsak *et al.*, 2015). The objectives of a CDM are, among others: to transform the data contained in the different databases into a common format (data model); to provide a common representation of these (terminology, vocabulary, codification schemes); and to carry out a systematic analysis using a written, standard package of analytical routines based on the common format.

The advantages of CDM stem from their use to create data repositories that use a unique metadata system and to transform the existing data into the standard schema of the CDM. The increased efficiency from this standardisation can speed up and streamline everything done with the data thereafter. Furthermore, CDM facilitate collaborative investigation given that they can generate evidence from a wide variety of sources.

Last, and notably, in RWD we can find different types of selection biases, confusion biases, information biases (also known as classification biases), etc. Like with all observation designs, it is very likely that RWD produces a selection bias. Some subjects have a greater probability of being observed than others, and therefore of being present in the sample. Selection bias is a systematic error that occurs when the distribution of frequencies of the explanatory variables is systematically different either among individuals or among groups (in particular among the intervention and control groups). Given that the result of interest is the result of the effect of the data processing after controlling the differences in the explanatory variables, a simple comparison of results among individuals and groups may not produce an unbiased estimation of the effect of the processing when the explanatory variables are distributed differently among these individuals or groups.

If the selection was exogenous, or in other words if the probability of a subject being observed was identical for all subjects, then weighting the sample to give

Background

less weight to the observed subjects would be sufficient. However, it is highly likely that the observed subjects are observed due to non-observed factors that intervene in this observation and would be correlated with the non-observable factors that affect the response variable. However, the probability that these subjects have been observed is not the same for all of them. Therefore, the weighting (standardisation) by age and sex would not correct the selection bias. In this case (called endogenous selection), more complex and suitable statistical methods have been used (two-part model, etc.).

The selection bias is the origin of the confusion bias. There is confusion bias when we observe a non-causal association between the exposure and the event under study or when we do not observe a real association between exposure and the event due to the action of a third variable that is not controlled (called a confounder or a confounding factor).

The errors that can occur in the measurement, in terms of both the exposure and the outcome, lead us to the information or classification bias. The errors to which we refer would consist in identifying the subjects that have not undergone the outcome as if they had (or have been exposed when they have not), and in assuming that some subjects have not undergone the outcome when in reality they have (or have not been exposed when they have). The information bias can be non-differential (random) when the degree of error is the same for the groups being compared, and differential (non-random) when the degree of error is different for the groups studied. The classification bias refers especially to this last type.

There are different methods to minimise these biases. Among the different methods to control the selection bias, we can highlight those classified as design methods and analysis methods, those that we could classify as methods based on observables, and methods based on non-observables. Obviously, all these methods are usually used in combination. Regarding the classification bias, the main methods proposed for their minimisation consist in establishing standardised proceedings for data collection, implementing measures/standard frameworks for measurement (per example, use of classification like CIE-10, ATP, etc.), and so on.

Background

1.5 EVENT DRIVEN DATA MODELS

RWD are obtained from multiple data sources. As a general rule, we can differentiate between the data sources that come from medical environments such as electronic health registers, clinical data (analyses, etc.), and invoices, and data that come from sources external to medical registers such as death registers, open data sources (Demski, 2016), data from smartphones, *wearables* (for example smartwatches), and other mobile devices (Reeder and David, 2016), in addition to data coming from sensors that make up the so-called Internet of Things (Scarpato et al., 2017). Among the open data sources are the open databases about air quality (Hasenkopf et al., 2015) and temperature, among others.

Some characteristics of RWD can be deduced in light of the previous definition. Since they come from various sources, the volume of data can be vast, especially if the sources are mobile devices. Compiling information from different data sources involves the need to manage multiple formats (JSON, XML, CSV, among others). This definition fits with the concept of complex data, defined as an information set that is not only numeric or symbolic, but is also multimedia and multistructured. The storing and processing of this type of information requires adapted structures (Boukraâ and Boussaid, 2009).

In epidemiology, a study of cohorts is a type of observational and analytical design (prospective or retrospective). What defines cohort studies is that the subjects are selected depending on the exposure of interest. In the simplest case, an exposed group and a non-exposed group are selected and the two are followed over time to compare the incidence of an event of interest. From one perspective of the use of an architecture based on events, we can think of a cohort study as an entity that reacts to different events produced in the exterior world (Fowler, 2006). This enables a cohort to be updated based on external source notifications.

It is within this context that the *eventr* package (Franquet *et al.*, 2020) of R (R Core Team, 2020) open code package is proposed, facilitating the implementation of architectures geared towards events (EDA).

Background

There are various advantages to structuring information using an EDA model: an audit trail, simplicity when it comes to debugging the code, the capacity to create alternate states, and the use of retroactive events (Fowler, 2017; Erb *et al.*, 2016).

- The audit trail is a register of all the events that have taken place concerning an object of interest. The events register when changes are made and when they happen.
- Debugging code errors is a simpler process because they are dealing with events that are independent of each other. It offers the possibility of creating testing environments where all the changes that have taken place are reproduced exactly, going backwards or forwards to the moment they occurred.
- Using events in a retroactive way enables the status of an object in a prior moment to be calculated more easily. This is especially useful in retrospective cohort studies.
- The architectures guided by events allows alternate statuses to be created. We can modify the definition of one of the events and re-calculate the status of an objects of interest retroactively.

Taking the qualities of this type of architecture into account, the creation and analysis of data based on RWD is simpler. Given the complexity and cost of the studies where this type of data is used, they are needed to be able to re-calculate the statuses of the individuals if there are coding errors, and to be able to audit all the steps that have to be taken to ensure the reproducibility of these studies. The creation of alternative statuses also facilitates the simulation of synthetic patients.

More details about the implementation that was carried out can be found in section 3.4 “Implementation of an R package for modelling data based on events”, and how this package can be used to model the data and using them in a simulated case and in a real case is explained in section 4.2.

Background

1.6 VISUALISATION OF INFORMATION

This section contains a description of what a visualisation is and what characteristics it must have. The terms visualisation and infographic will be used as synonyms.

There are different definitions of infographic, but the definition that will be used throughout this thesis is the one proposed by Cairo (2016): “A visualisation is any visual representation of information designed to communicate, analyse, discover, and explore the data”. This definition includes the traditional concept of infographic found in newspapers and magazines, and the visualisations carried out by other means, such as via browsers and smartphones.

1.6.1 Characteristics of a visualisation

To fulfil their purpose, an infographic must be truthful, functional, beautiful, insightful, and enlightening.

An infographic must be truthful. This means that the data contained in it and the way in which it is presented must be as close to representing reality as possible. An infographic containing invented data or data that does not represent the reality does not represent information and so cannot be considered an infographic.

Given that all infographics must have an objective, fulfilling this objective is key. Therefore, to be able to do so an infographic must be functional. If we want to explain what climate change is and what its effects are, the infographic must ensure that the users are able to understand this phenomenon, thanks to the infographic.

An infographic must be visually beautiful. One of the functions of all infographics, apart from informing and showing information, must be to attract users' attention. People's curiosity about the infographic must be peaked so that they will want to read it.

One of the advantages of using graphics to represent information is their capacity to illustrate patterns and characteristics in a simplified way. This quality is called being insightful.

Background

If the four characteristics described above are present, a good infographic will also be enlightening. This means that like a technical book an infographic must enable their objective public to find out something that they did not know, such that it can change the way they see the world in relation to the subject in question.

1.6.2 Methodology in the process of creating an infographic.

The working methodology used to develop the infographics is described below. It is the one proposed by the Government of Catalonia in the “Data Infographic Guide” (Government of Catalonia, 2018). This methodology proposes that the work be divided into three phases made up of various tasks. The three phases are defining the strategy, the recompilation and transformation of the data, and the design of the infographic.

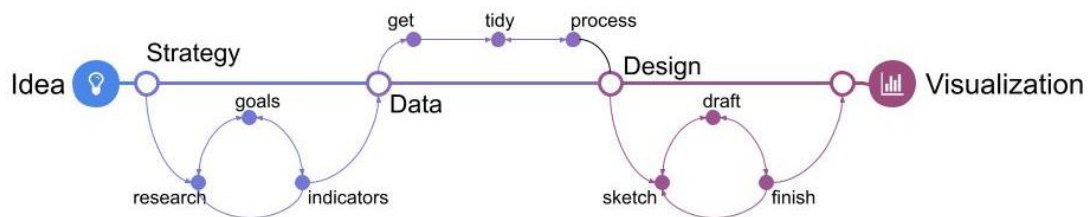


Figure 4 Model of the process of creating infographic. Source: Author's own elaboration.

1.6.2.1 Strategy

The strategy consists in defining the subject and the objectives of the infographic, and to do so a bibliographic search of the subject must be carried out. This must be used to define the scope of the infographic, helping us to pose the appropriate questions and provide answers to them. It is recommended to speak to experts about the subject to be developed wherever possible.

The questions to which we want to find answers should be able to be defined from the bibliographic research and consultation with experts. The indicators that allow us to find the answers are selected based on these initial questions.

The “Guia de visualització de dades” (Generalitat de Catalunya, 2018) proposes using the 5W1H scheme. This tool used in journalism consists in answering the

Background

questions what has happened, who is responsible for what has happened, where did it happen, when did it happen, why did it happen, how did it happen?

1.6.2.2 *Data*

Once the objective has been defined and detailed based on a set of questions to answer and their corresponding indicators, the next step is to research, transform, and exploit the data.

The data can be obtained from public data and other information sources such as the databases of one or several entities. The infographic “La demografia a l’Alt Empordà”, explained in more detail in section 4.1.1.1, is an example of an infographic created based on public data, in this case from Idescat. The infographic “La Natalitat a l’Alt Empordà”, explained in section 4.1.1.2 and available in the Appendix, entitled La Natalitat a l’Alt Empordà, and on the Indika website, is an example of where public data provided by Idescat and Fundació Salut Empordà have been used.

The most frequently used formats are CSV, JSON and XML. To exploit these data, they need to be transformed into a structure that is more adapted to our needs. The R programming language (R Core Team, 2020) and the set of tidyverse packages (Wickham *et al.*, 2019) were used to create the infographics in this thesis.

1.6.2.3 *Design*

The design is the last part of the process and consists in actually creating the infographic. An iterative methodology consisting in drafting, prototyping, and finalizing was used in this thesis to do so.

Drafting is a process whose objective is to allow the designer to discover what form the infographic should take, what graphics could be used, and where each graphic should be put so that the infographic is coherent. A draft can be a paper

Background

or any other tool that can be used to capture the idea on a medium that is easy to share and can be produced quickly.

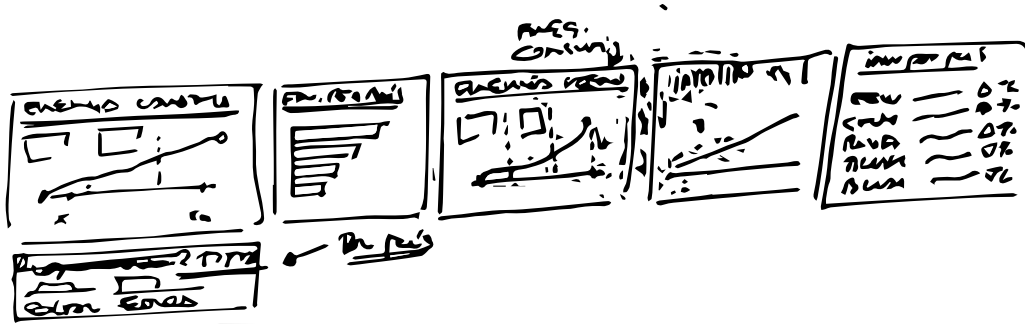


Figure 5 Sketch of an infographic. Source: Author's own elaboration.

Once a draft has been produced and there is a general idea of what infographics should be like, a prototype is constructed using appropriate graphic and IT tools. A combination of (R Core Team, 2020) and the package ggplot2 (Wickham, 2016) to produce the graphics in SVG format were used in the design of the infographics in this thesis. Once the graphs were obtained, the design was produced using Adobe Illustrator (Adobe Inc., 2019) and *Inkscape* (Inkscape Project, 2020). The objective was to share the prototype with the experts involved in the creation process.

Last, the texts must be reviewed and the functional infographics approved. In this phase, it is recommended to look for a profile of person in line with that of the objective public to confirm that their understanding is the hoped for one. The “Guia de visualització de dades” (Generalitat de Catalunya, 2018) recommends a total of five people for each iteration because this is how the most relevant changes to be implemented can be found without the need to make excessive use of the resources.

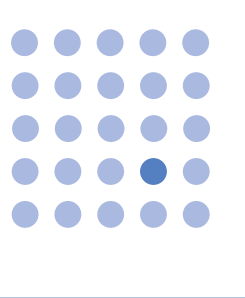
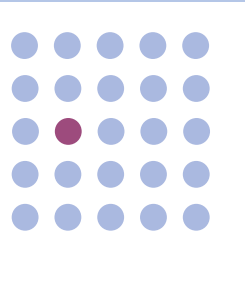
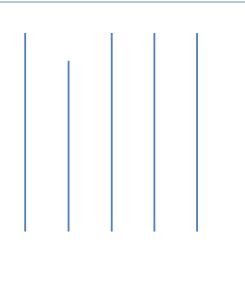
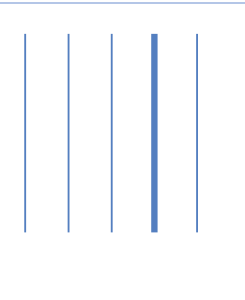
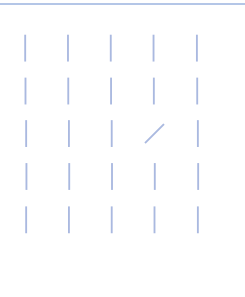
1.6.3 Principles of an information infographic

In this section, some of the fundamental principles of the information infographic are described from the neurological point of view. The objective is to explain how the brain interprets the information it receives via the sense of sight. Few (2013)

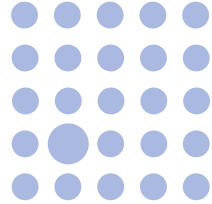
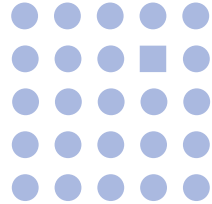
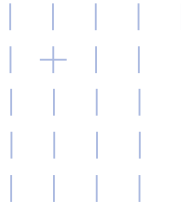
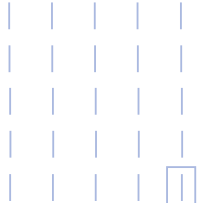

Background

selects eleven attributes of visual perception grouped in four large categories: colour, shape, position, and movement.

Table 2 Attention attributes prior to visual perception

Category	Attribute	Illustration
Colour	Intensity	
	Hue	
Form	Length	
	Width	
	Orientation	

Background

	Size	
	Shape	
	Marks	
	Enclosure	
Position	2D position	
Movement	Flickering	The cursor of the mouse in the latest text editors flickers to show

Background

		where in the text we are.
--	--	---------------------------

Source: Author's own elaboration based on Few (2013).

1.7 OPEN DATA

La *Open Knowledge Foundation* (OKF) defines open data as the data that can be used, reused, and freely redistributed by anybody, provided that sufficient credit is given and they are distributed in the same way as they have been published (Open Knowledge Foundation, 2021a).

There are three criteria that must be fulfilled for data to be considered as “open”: (1) The criteria of availability and access; (2) the criteria of reuse and redistribution; and (3) the criteria of universal participation.

The first criteria guarantees that the information is available as a set and at a reasonable reproduction cost, preferably downloadable via the Internet. Furthermore, the information must be in a form that is convenient for its use.

The second criteria guarantees that the data allow for their reuse and redistribution, and that they can be integrated into other datasets.

Last, the third criteria guarantees that everybody can use the data without any type of discrimination to groups or individuals. Commercialisation of the data must be restricted to ensure that other people do not have to pay for a resource that is open. The use that can be made of the data must be specified by means of a specific license.

Open data must allow interoperability with other data, entities, people, and agents. Interoperability is the capacity to integrate different data sets and in terms of open data it is the capacity to be combined with other data and develop better products and services.

Background

Open data must be of a non-personal nature. It must not contain information about specific individuals. Other types of data, such as those that are important for national security, must be protected or be subject to special restrictions.

In the process of opening the data, the institution does not have to open all the data at once. It is best to open a small, easily compilable database.

Once the data is open, it is important to communicate with the users and people who will make use of it. Users' needs must be aware of what other datasets can be of use to them.

It must be remembered that a large proportion of the data published does not reach the users directly, but through intermediary agents who transform these data into other products such as infographics, websites, or other types of support.

The Open Knowledge Foundation (OKF) defines four phases in its data opening process: (1) Select the data to be opened, (2) select an open license that details the intellectual property rights of the information, (3) transform the data into an easily transferable and exploitable format such as JSON or XML, and (4) make them visible, publish them on the entity's website, social networks, or a central catalogue of databases.

Depending on the size of the institution that wants to offer data, the cost of the selection process can be very high due to the complexity of the data stored. In these cases, it is recommended to make a list of the data available and to later compare this to another list compiled following consultation with the community and those who will use the data to select those that most satisfy user demand.

When publishing the information, it is important to decide what type of license will be used. Table 3 shows a list of licenses that can be used when datasets are opened and that fulfil the definition of "open", according to the OKF.

Table 3 Licenses that are suitable for open data.

Name of the license	Requires attribution	Require share-alike
---------------------	----------------------	---------------------

Background

Creative Commons CCZero (CC0-1.0)	No	No
Open Data Commons Public Domain Dedication and Licence (PDDL-1.0)	No	No
Creative Commons Attribution 4.0 (CC-BY-4.0)	Yes	No
Open Data Commons Attribution License (ODC-By-1.0)	Yes	No
Creative Commons Attribution Share-Alike 4.0 (CC-BY-SA-4.0)	Yes	Yes
Open Data Commons Open Database License (ODbL-1.0)	Yes	Yes

Source: Author's own elaboration based on the Open Knowledge Foundation - Commformant Licenses

The data must be open legally and technically to be considered open. Being open technically means that they are available in a machine legible format.

1.7.1 Open data classification system

Tim Berners-Lee describes an open data development system based on the classification of data using a system of stars. Table 4 shows the stars classification described on the 5stardata website (<https://5stardata.info/en/>). The column headed "Number of stars" indicates the number of starts that would be assigned to a data set that meets the requirements set out in the "Description" column.

Table 4 Classification of open data using a system of stars.

Number of stars	Description
-----------------	-------------

Background

1	Make your stuff available on the Web (whatever format) under an open license.
2	Make it available as structured data (e.g., Excel instead of image scan of a table)
3	Make it available in a non-proprietary open format (e.g., CSV instead of Excel)
4	Use URIs to denote things, so that people can point at your stuff
5	Link your data to other data to provide context (like, for example RDFa).

Source: <https://5stardata.info/en/>

1.7.2 JavaScript Object Notation

JavaScript Object Notation (JSON) (ECMA International, 2017) is a lightweight data-interchange format based on text, irrespective of the language, to define data exchange formats. It was derived from the ECMAScript programming language (ECMA International, 2017), but it is independent of the programming language. JSON defines a small set of structuring rules for the portable structuring of structured data.

JSON syntax is based on two basic structures: (1) A collection of number/value pairs (2) an ordered list of values.

The objects are defined using brackets, as shown in Figure 6.

Background

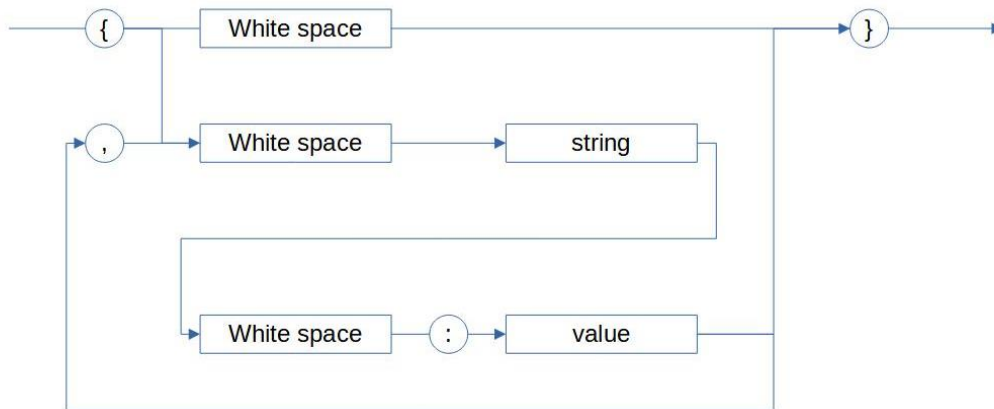


Figure 6 JSON object schema. Source: Author's own elaboration based on <https://www.json.org/json-en.html>

```
{
  "firstName": "Alvaro",
  "secondName": "Franquet"
}
```

A matrix is an ordered collection of values. A matrix starts with the left bracket ([) and ends with a right bracket (]). Between brackets, there are numbers separated by commas. Figure 7 shows the representation of a matrix.



Figure 7 Definition schema of a matrix using JSON. Source: Author's own elaboration based on <https://www.json.org/json-en.html>

```
[1, 2, 3, 4]
```

Background

A value can be a *string* type between inverted commas, a number, a boolean value (*true* or *false*), a null value, an object, or a matrix. These structures can overlap.

1.7.3 Extensible Markup Language (XML)

Extensible Markup Language (XML) is a syntax defined by the *World Wide Web Consortium (W3C)* (Extensible Markup Language (XML) 1.0 (Fifth Edition), 2008) designed to store and transfer data using a network with the aim of ensuring it is legible for both humans and machines.

Like the language HTML, XML uses labels to define structures contained in the document which, unlike the first, have not been previously defined, or in other words it is the author of the XML document who defines their own labels.

A tag is a chain of characters contained between the less than (<) and the more than symbols (>). Figure 8 shows the schema related to the definition of these tags.

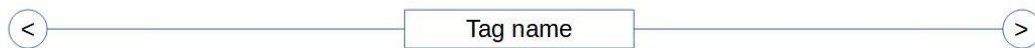


Figure 8 Schema of a label in XML format. Source: Author's own elaboration.

```
<firstName>
```

An element is a set made up of two tags, one that indicates when the element starts, and one that indicates when the element ends, characterised by the use of the forwards slash (/) followed by the name of the tag. Figure 9 shows the general schema to create an element.

Background



Figure 9 Schema of an element in XML format. Source: Author's own elaboration.

```
<person>  
  <firstName> Álvaro </firstName>  
  <secondName> Franquet </secondName>  
</person >
```

The elements can contain attributes, which must go between inverted commas and can contain the same information. This is why there are no clear rules on the use of one or the other in each situation. If there are some characteristics that can help the author of a document to decide on this matter, they are: (1) the attributes can only have one value, while the elements can have more than one, (2) the attributes cannot interleave, whereas the elements can, and (3) the attributes are more difficult to expand. Figure 10 shows the basic schema of an element with attributes.

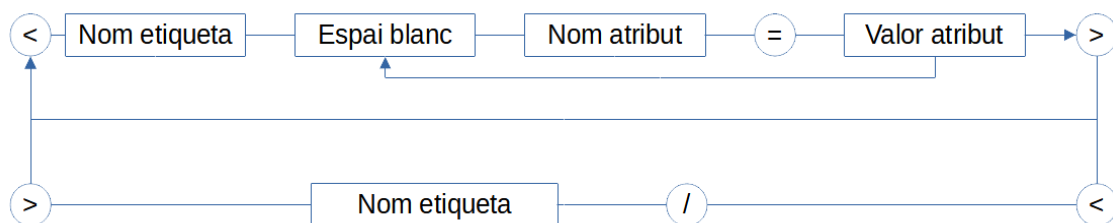


Figure 10 Schema of an element in XML format with attributes. Source: Author's own elaboration.

Background

```
<person sex = "male">  
  <firstName> Álvaro </firstName>  
  <secondName> Franquet </secondName>  
</person >
```

2 HYPOTHESIS AND OBJECTIVES

2.1 HYPOTHESIS

Main hypothesis

- It is possible to integrate data associated with the health observatory and *real world data*.
- The exploitation of this integrated information can improve the state of well-being in the county.

Secondary hypotheses

- The application of suitable statistical methods will allow the improvement of quality and will prevent biases, which in turn will enable the internal validity of the database to be achieved.
- The external validity of this database will be achieved via the reproducibility.

2.2 OBJECTIVES

The Indika Health and Social Pole of the Alt Empordà was created with the aim of facilitating the generation of knowledge of the impact of social determinants on health. Its purpose is to be of assistance when formulating effective health policies via the use of health indicators as the project's main tool. The use of these indicators must serve to guide intersectoral strategies and align the health and social agents to work in a coordinated way.

General objectives

- Achieve a scientific-technical health observatory unit that will promote the situation of health and its determinant factors, and especially the socioeconomic ones, in Alt Empordà.

Hypothesis and objectives

- Supply the public and private entities that can intervene in improving the state of well-being in the county of Alt Empordà with information that comes from integrating the health observatory and *Real World Data*.

Specific objectives

- Construct and maintain a repository of health, socioeconomic, and environmental information in a centralised place that enables the storage, maintenance and dissemination of digital health, socioeconomic, and environmental data related to Alt Empordà, and which are also relevant at the levels of Catalonia, Spain, and Europe.
- Design a database of health, socioeconomic, and environmental indicators composed of:
 - Health results: mortality, morbidity (objective and self-perceived).
 - Health determinants: quality of care, lifestyles.
 - Socioeconomic factors.
 - Environmental quality.
- Create and manage a database using so-called *Real World Data*, or in other words, secondary information coming from primary and specialised health care (including hospitalisations and emergencies) in the county of Alt Empordà.
- Develop statistical methods to control the quality of this database.
- To integrate this *Real World Data* database and the health observatory.
- To exploit the integrated indicators database, generating descriptive reports (tables and graphics, crossed, prevalence and incidence tables, standardised rates), maps (graphical representation of the standardised rates, socioeconomic factors, environmental), and other technical documents.

3 METHODS

3.1 DESIGN A REPOSITORY OF HEALTH, SOCIOECONOMIC AND ENVIRONMENTAL INDICATORS

This section contains the procedure that was used to construct the information repository with health, socioeconomic and environmental indicators. It is divided into three subsections. In the first, the procedure used to define the requirements of the repository is described, and the second explains the design used to construct it and how this was done.

3.1.1 Identification of requirements

A repository is a computer system where an organisation's information is stored so that its members (be they people or entities) can share it in an organised and standardised way. The Indika data repository provides access to indicators of interest about the territory of Alt Empordà.

The identification of requirements is the process by which the functional relations are identified. These are understood to be the set of characteristics that make up the global product, with the intention of identifying its sub-components. In short, a requirements identification process is a detailed breakdown of the structures that make will make up the global system to be able to define the objectives of each of the components. The intention is to divide a system in such a way that each component can be described without the need to refer to another of the components. This way, each part of the system will have independent functions that can be reused and replaced.

The objective of the repository is to facilitate the process of reusing relevant open data concerning the territory by its various agents. The strategic guide for launching minimum datasets for publication (Federación española de municipios y provincias, Red de entidades locales por la transparencia y participación ciudadana, Wolter Kluwer, 2018) defines open data as data with a license that allows their reuse.

Methods

Among other objectives, the aim of the Health and Social Pole of the Alt Empordà is to create a network of entities and organisations which work together to improve the situation of the county in a social and health sense. Therefore, one of the requirements that was kept in mind was the need for collaboration among the different bodies, ones of which is the Alt Empordà Social Observatory and the Fundació Salut Empordà.

The following general requirements were decided based on discussions with those involved in Indika and on the bibliographic review of multiple observatories and repositories: (1) the data must be open, (2) given the initial resources available to Indika, the need for them must be reduced by means of facilitating collaboration among entities, (3) it must facilitate scientific production, (4) the data must be self-contained, meaning that it can be used as an individual item, (5) the data must be able to be grouped in specific topics of interest, and (6) the data must be open access.

Open data are data that are accessible and reusable without any specific permissions. The entity *Open Knowledge Foundation* defines open data as data that can be freely used, reused, and redistributed among anybody and are subject to the requirement of attributing and sharing in the way in which they appear, provided the security and privacy of the information is respected at all times.

The process to identify the requirements was based on the review of the set of indicators presented in the report *Key Figures on Europe – Statistics Illustrated* published annually by EUROSTAT, the *Marco de indicadores mundiales para los Objetivos de Desarrollo Sostenible y metas de la Agenda 2030 para el Desarrollo Sostenible* (ONU, 2020), and the European Core Health Indicators shortlist of 88 health indicators identified by policy area (European Commission).

3.1.1.1 Review of indicators

The review of indicators consisted in the search for available indicators for the county of Alt Empordà and, wherever possible, for its municipalities. The indicators described in *Key figures on Europe — Statistics illustrated — 2019 edition* (EUROSTAT, 2019) were used to carry out the search.

Methods

A Table with the code, name, and description of the indicator was created. The availability of the each of the indicators at the European, Spanish, Catalan, Provincial, County and Municipal levels was recorded for each of them. According to the desired level of breakdown, the EUROSTAT, INE or IDESCAT websites were consulted first.

Once the indicators that were not publicly available were identified, Idescat was consulted. Following this consultation, the information was requested from the organisations responsible for it (according to the response of Idescat).

3.1.1.2 Technical requirements

The technical requirements that must be met by the repository are described in this section. It must be an information store promoted by the community. Taking Indika's philosophy into account, it must be adapted to the needs of the territory's various agents. These agents must therefore be capable of deciding what they want this repository to be like and what information it must contain, and they must also take part in collecting and updating the data.

Given the nature of the observatory, it is essential that its repository is decentralised. This means that it is not maintained by one single person, but that all the people and organisations involved are responsible for maintaining it and keeping it up to date. To this effect, if an organisation want to stop taking part in the project, the rest can continue with it based on the work carried out. Similarly, if a company or an organisation wants to use it on a local scale, they must be capable of downloading the repository and using it without the need to be connected to the network.

The repository must be capable of including new indicators in the future without any difficulty. It must be easily expandable. To meet this objective, the repository must be made up of small files, each of which contains information related to an indicator in a way that it is extremely easy to include a new indicator.

We would like the organisations taking part to receive due credit, which is why the repository must record who sends in each indicator and who is responsible for its updating. In other words, the traceability of the information must be

Methods

assured.

The repository must be open. The data included in it must be accessible for everyone and the modification of the information must be allowed to guarantee its reliability. Given the capacity to monitor any updates (traceability), it is easy to give each contributor due credit.

3.1.2 Design and architecture of the repository

Having identified the minimum requirements of the repository, we proceeded to design its architecture, taking the needs of the organisation Indika and its external and internal characteristics into account.

3.2 CONSTRUCTION OF A HEALTH, SOCIOECONOMIC AND ENVIRONMENTAL REPOSITORY FOR ALT EMPORDÀ

Once the observatory's requirements had been established, their implementation was effected as a GitHub repository (GitHub, 2020). GitHub is a collaborative development platform that allows open and closed code projects to be accommodated with a versions control system. A versions control system serves to register the changes made to a file over time for the purpose of being able to recover a specific version later on. GitHub used the Git versions control system (Chacon and Straub, 2014).

The repository is a GitHub project with a GPL-3 license (Free Software Foundation, 2020). The GPL-3 is a type of software license that allows code to be copied, distributed (commercially or not), and modified, on the proviso that any modification continues to be distributed with the same GPL license. The GPL license does allow for the distribution of executable programmes without the corresponding source code or an offer of how to obtain it for free. It is the most used free software license.

The repository is divided into two directories: *indicators* and *indicators-list*. The *indicators* directory contains the data and descriptors of the indicators. Each indicator is found in a file in JSON format. The *indicators-list* contains lists of indicators. The lists of indicators are files in JSON format that group a set of indicators according to their theme or interest. For example, the list of indicators

Methods

01-indicadors-clau-de-el-alt-emporda.json contains the list of indicators used in the “Key Indicators of Alt Empordà” document (Franquet *et al.*, 2020). The function of the lists of indicators is to facilitate the download of a set of indicators and to share groups of indicators by themes of interest.

The indicators are divided into two attributes: meta and data. The meta attribute is an object with six attributes that describes the information contained in the data. Table 5 shows the attributes that meta information must contain.

Table 5 Meta information of the indicators on the repository

Attribute	Description	Type
name	The name of the indicator. It must be a name identical to the name of the file	character
Title	The name of the indicator	character
Description	The description of the indicator	character
Source	The source from which the data have been extracted. Wherever possible, a direct link	character
maintainer	The person responsible for maintaining the indicator in the repository	character

Methods

lastUpdate	Date the indicator was last updated in the repository.	Date and time
------------	--	---------------

Source: Author's own elaboration

3.2.1 API

An applications programming interface (API) is a description of the communication protocol between two agents (normally a machine and a human or two machines). The function of an API is to facilitate and standardise the form in which the agents relate to the software in question. An example of API are the R programming packages, which facilitate the application of statistical methodologies using specific functions.

3.3 USE OF THE INTEGRATED DATA BASES

3.3.1 Key Indicators of Alt Empordà

The document “Key Indicators of Alt Empordà” is a collection of indicators of different topics such as the population, the economy, and the environment in Alt Empordà. This will be an annually produced document, the objective of which is to produce a global vision of the situation of Alt Empordà. The document is structured following the EUROSTAT document “*Key Figures on Europe – Statistics Illustrated 2019 version*” (EUROSTAT, 2019). The first step in compiling this document was to review the indicators provided by EUROSTAT and to evaluate their availability at different territorial levels, including Spain, Catalonia, the province of Girona and Alt Empordà.

3.4 IMPLEMENTATION OF AN R PACKAGE FOR MODELLING DATA BASED ON EVENTS

This section is divided into two sub-parts. The first describes what an architecture based on events is and what parts an architecture based on events has. In the second part, the implementation carried out in the eventr package is described.

Methods

An architecture or a data model is an abstract model about what information describes a set of data and how this information is organised in a detailed way (Simsion i Graham, 2004).

An event driven architecture (EDA) makes use of the events to trigger a set of actions. In its simplest form, it has three key elements: the event, the handler, and the dispatcher. The event is responsible for transporting the information needed to carry out an action. The handler will process the event information to execute a task. The dispatcher will evaluate the events and decide which handler must be used depending on the type of events. Figure 11 shows the flux diagram.

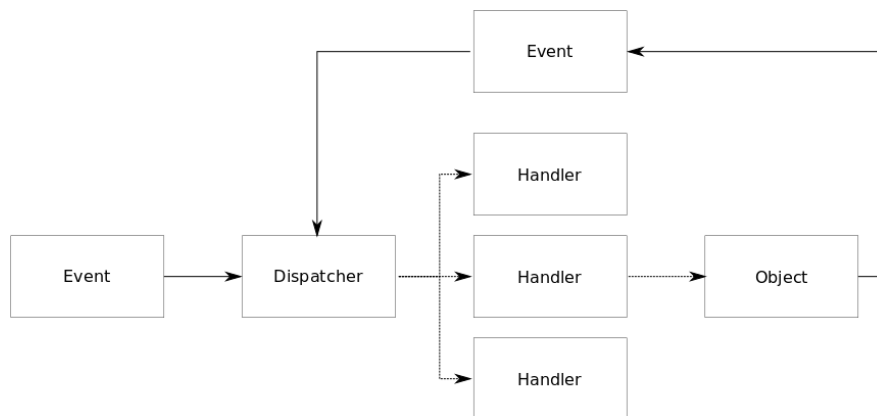


Figure 11: Flux diagram of an event driven architecture. Source: Author's own elaboration

3.4.1 Terminology

Below is a description of the terminology used:

- The events are objects with a with a header and a body. The header stores all the information related to the same event such as, for example, its type (`type`), a time marker (`time`) and an identifier (`id`). The body stores the information that describes what has happened during this event.
- The specification of an event is the definition of its structure. The specification indicates what information it contains and what this structure must be like.
- The occurrence of an event is each one of the occurrences of a specification that have taken place.

Methods

By way of example, let's assume we are interested in defining the event Birth. The specification of the event Birth defines this event as an object that has three parameters: an identifier (`id`), the date of birth (`birth_date`) and the place of birth (`birth_place`). The occurrence of the event Birth will be each of the births in our registry. Table 6 shows examples of the occurrences of the event Birth.

Table 6 Example of occurrences of three events. Source: Author's own elaboration.

id	birth_date	birth_place
eduardo-punset	1936-11-09	Barcelona
serj-tankian	1967-08-21	Beirut
ara-malikian	1968-09-14	Beirut

- The Handlers are responsible for applying the processes needed to transform their parameters into the attributes of a project.
- The Dispatchers are responsible for sending events to their respective handlers.

3.4.2 Implementation

Below is a description of how the eventr package was implemented. Three classic principles have been included in this implementation: `event`, `handler`, and `dispatcher`. Following the model proposed by Wickham (2019), each class has a constructor, a validator, and a helper.

The class `event` is a list made up of two elements: a header and a body. The header stores the information related to the event, which is an identifier of the event (`id`), time (`time`) when it took place, and the type (`type`). The element `body` stores the information that will serve to transform the attributes of an object of interest.

To facilitate the manipulation and use of the events, the following functions have been implemented.

Methods

- `get_body()`: Returns the list of elements stored in the body of the event.
- `get_body_attr()`: Unlike the `get_body()` function, it only returns the attribute that coincides with the parameter `attr`.
- `get_header()`: Returns a list with information about the header of the event.
- `get_id()` , `get_time()` and `get_type()`: Returns the identifier, the time of the event, and the type of event, respectively.

The events can be grouped into a list of events `event_list`. The list of events can be created using the function `event_list()` or by means of the operator `+`. This is how a list of events and the sum of all the events can be defined.

```
R> my_events_list <- first_event + second_event + nth_event
```

The function `handler()` returns a `handler` type object. The `handlers` are responsible for modifying the state of an object based on the information contained in an event. The function `handler()` has got two parameters: `type` and `FUN`. The first serves to indicate what type of events the `handler` must manipulate. `FUN` is the function responsible for modifying the state of an object based on the information contained in the event. The function `FUN` must be of the type:

```
R> foo = function(obj, event) {  
  obj <- do_something(event)  
  return(obj)  
}
```

Where `obj` is the object to modify and the `event` is the input event. In other words, the function must have an object (`obj`) and an event (`event`) as input parameters, and it must return only a modified object (`obj`).

The following functions have been implemented for the class `handler`:

Methods

- `get_type()`: Indicates the type of event the handler is for.
- `get_fun()`: Returns the function (FUN) of an object type handler.
- Like the events, the handlers can also be grouped in lists. The function `handlers_list()` is responsible for creating them. The `handlers_list` can only contain a handler for types of event. If we introduce a handler for a type of event that already has one assigned to it, the second one overwrites the first one.

```
first_handler <- handler(  
  type = 'FIRST_EVENT_TYPE',  
  FUN = function(obj, event) return(obj)  
)  
second_handler <- handler(  
  type = 'SECOND_EVENT_TYPE',  
  FUN = function(obj, event) return(obj)  
)  
  
my_handlers_list <- handlers_list(first_handler, second_handler)
```

It is also possible to use the operator `+` to create a list of handlers:

```
my_handlers_list <- first_handler + second_handler
```

The class `dispatcher` is a list made up of a `handlers` and the function `dispatch()`. Given a list of events, the function `dispatch()` distributes these events between their corresponding `handlers` in a sequential way. The function `dispatch()` has got three input parameters: `obj`, `events` and `accumulate`.

Methods

`obj` is the object of interest, which will be modified by the events. `events` is a list of events and `accumulate` is a boolean input that indicates whether the output function must be the final state of the object or the state after each one of the events included in the list of events (`events`).

For the class `dispatcher`, the following methods have been implemented:

- `add_handler()` o +: Allows a handler and a dispatcher to be added.
- `get_handlers()`: Allows the list of handlers included in a dispatcher to be recovered.

`get_dispatch()`: Returns the `dispatch()` functions created from the handlers.

4 RESULTS

In this chapter, the results obtained from the activity carried out within the framework of this thesis are presented. The chapter is divided in two parts. The first part corresponds to the analysis of the impact on the territory the observatory has had, and the second part corresponds to the analysis of the impact that the eventr package has had.

4.1 INDIKA HEALTH AND SOCIAL POLE OF THE ALT EMPORDÀ

One of the main objectives of the project was to evaluate the capacity to exploit the integrated information to improve the state of well-being in the county. One of the indicators used to evaluate this capacity is the number of projects drawn up. A total of 5 infographics were created, plus 1 web application on the COVID-19 situation in Alt Empordà, the report "Key Indicators of Alt Empordà" (Franquet *et al.*, 2020), and the Indika website (Indika Pol de Salut i Social de l'Alt Empordà, 2020).

4.1.1 Infographics

In this section, the three infographics created from the observatory are described: Població a l'Alt Empordà (*Population in Alt Empordà*), La natalitat a l'Alt Empordà (the *natality in Alt Empordà*) i Un any de COVID-19 a l'Alt Empordà (*Un any de COVID-19 a l'Alt Empordà*).

4.1.1.1 *Population in Alt Empordà*

The first infographic created was "Població a l'Alt Empordà" (Fundació Salut Empordà, 2019). The aim of the strategy adopted was to answer the questions: What is the population of Alt Empordà? What is the child population? What is the population aged 64 years and over? What is the over aged population? The latter is considered as those over 84 years old. The infographic aims to delimit the data in its geographic context, making comparison with the information available for the whole of Catalonia possible.

Results

The data used for the infographic come from the statistics: “Structure by age, ageing, and dependency” (Idescat, 2020a) and “Population on 1 January. By sex and age year on year” (Idescat, 2021a).

The transformation and exploitation of the data was carried out using the R programming language (R Core Team, 2020). The corresponding code can be found in the appendix of this thesis named *Code of the “Population in Alt Empordà”*.

Once the data was transformed and a descriptive analysis of them made, various drafts were prepared, as shown in Figure 12.

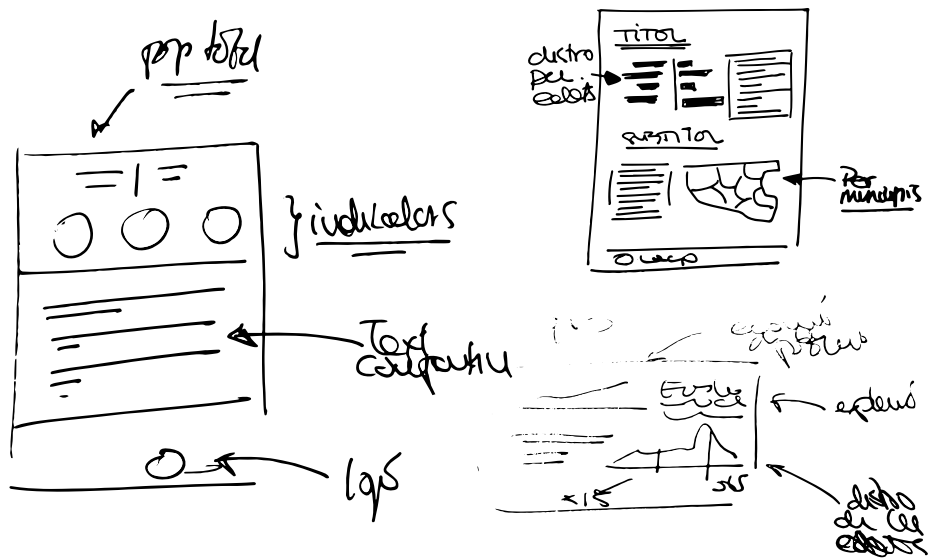


Figure 12 Three drafts for the “Population in Alt Empordà” Infographic. Author’s own elaboration.

Once the draft was produced and which of the options is most fitting decided, a prototype with the real data was created and the text and the different explanations revised. Figure 13 shows the final result, which can also be found on the Indika website (Fundació Salut Empordà, 2019) and as the Infographic in the Appendix.



Figure 13 " Població a l'Alt Empordà " Infographic version 2019. Author's own elaboration.

4.1.1.2 The "La Natalitat a l'Alt Empordà " Infographic

The objective of the "La Natalitat a l'Alt Empordà" Infographic was to show the information related to the births in 2020. The birth rate of a region is one of the key demographic growth values and so its analysis is important.

The objective of the infographic was to answer the following questions. What has happened with the birth rate in Alt Empordà in terms of its evolution? What is the current structure of the birth rate and why does it have this structure? And in relation to the SDG, is the Alt Empordà birth rate sustainable and healthy?

The data used to create the infographics was obtained from the Fundació Salut Empordà registries, from the gross birth rate (Idescat, 2021b), and the births by mothers' age (Idescat, 2021c). The code with which the data were explored and transformed can be found in the appendix, named Code of the Code of the "La Natalitat a l'Alt Empordà" Infographic.

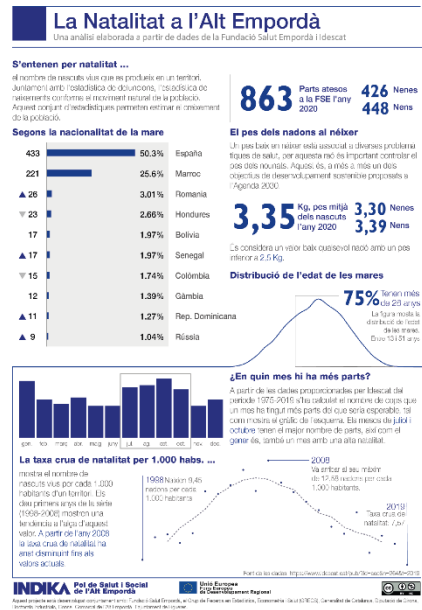
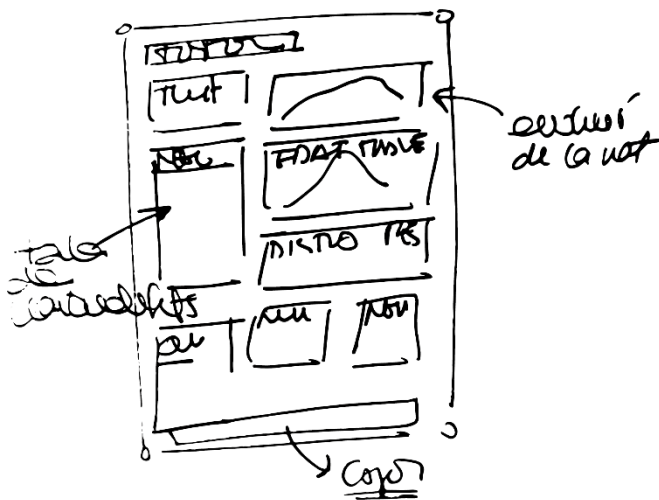


Figure 14 Draft of the infographic “Birth rate in Alt Empordà” and the final infographic. Author’s own elaboration.

Figure 14 shows one of the drafts produced and the final “Birth rate in Alt Empordà” infographic, which can also be found in the La Natalitat a l’Alt Empordà on the Indika webpage (Indika Health and Social Pole of the Alt Empordà, 2021).

4.1.1.3 “Un any de COVID-19 a l’Alt Empordà” Infographic

Due to the situation caused by the COVID-19 pandemic, and given the objective of the observatory, different actions were carried out with the aim of informing the population about the state of the county in this respect. More specifically, this was done via a web application, which is described in section 4.1.2, and by means of the “Un any de COVID-19 a l’Alt Empordà” infographic, which analyses how the pandemic has behaved over the year in greater depth.

The objective of the infographic was to show the evolution of the pandemic in the county and whether there have been any differences between its behaviour in Alt Empordà and in the rest of Catalonia, as well as the differences in the situation of the vaccination programme.

Different information sources were used to create the infographic: “Registry of COVID-19 cases in Catalonia. Segregation by sex and municipality” (Departament de Salut de la Generalitat de Catalunya, 2021a)[Department of Health of the Government of Catalonia, 2021a], Registry of COVID-19 cases in

Results

Catalonia. Segregation by sex and age” (Departament de Salut de la Generalitat de Catalunya, 2021b), the “Registry of deaths due to COVID-19 in Catalonia. Segregation by sex and county” (Departament de Salut de la Generalitat de Catalunya, 2021c), “Vaccination against COVID-19 by county” (Departament de Salut de la Generalitat de Catalunya, 2021d), and the number of hospital admissions due to COVID-19 in the FSE.

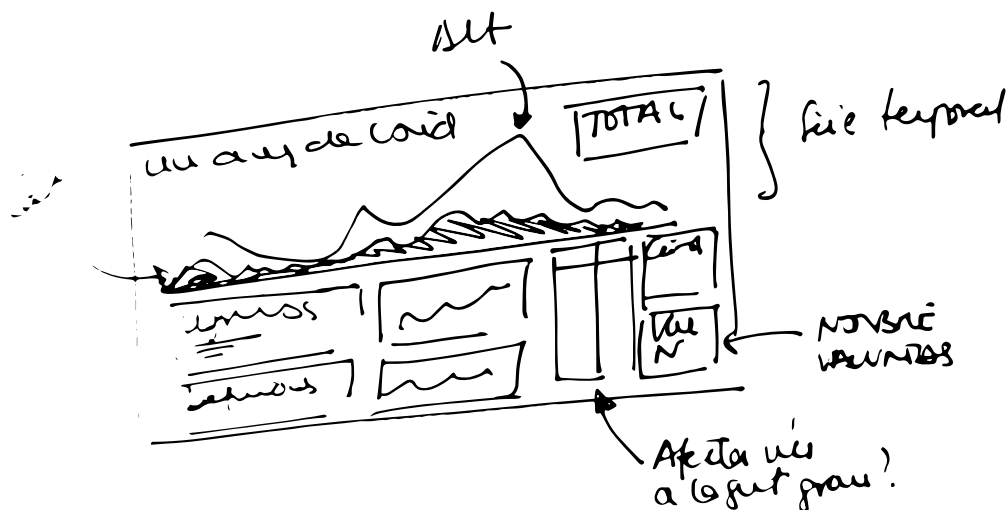


Figure 15 Draft of the infographic "Un any de COVID-19 a l'Alt Empordà". Source: Author's own elaboration.

Figure 15 shows the draft for the design of the infographic and Figure 16 shows one of the first prototypes from which it was decided to modify its structure to make it cover two pages.

Un any de COVID-19 a l'Alt Empordà

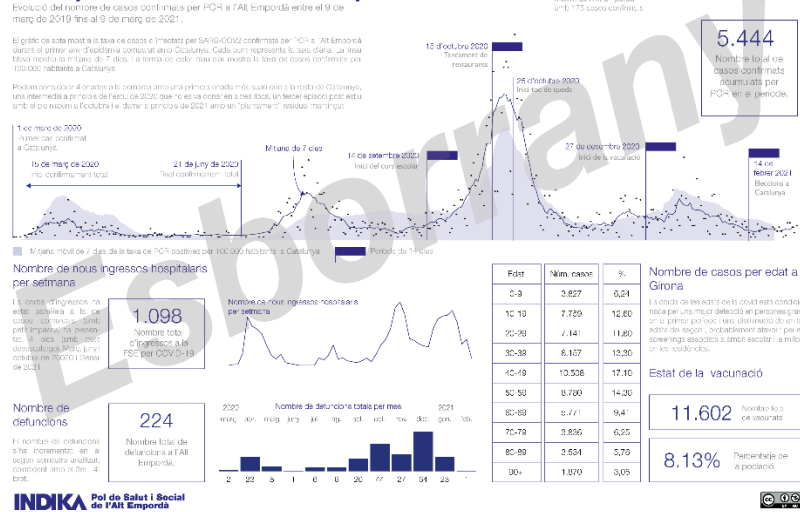


Figure 16 Prototype of the infographic "Un any de COVID-19 a l'Alt Empordà". Source: Author's own elaboration.

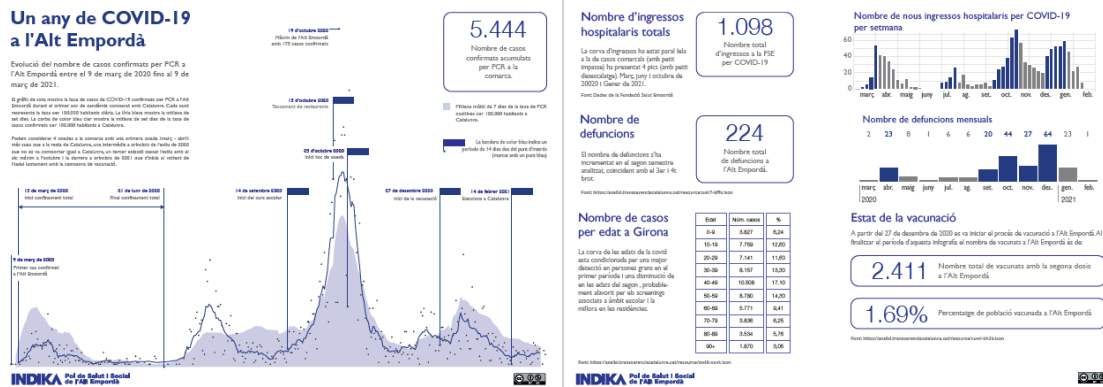


Figure 17 Infographic "Un any de COVID-19 a l'Alt Empordà". Source: Author's own elaboration.

Figure 17 shows the final infographic available in the appendix , together with the R code used for its elaboration.

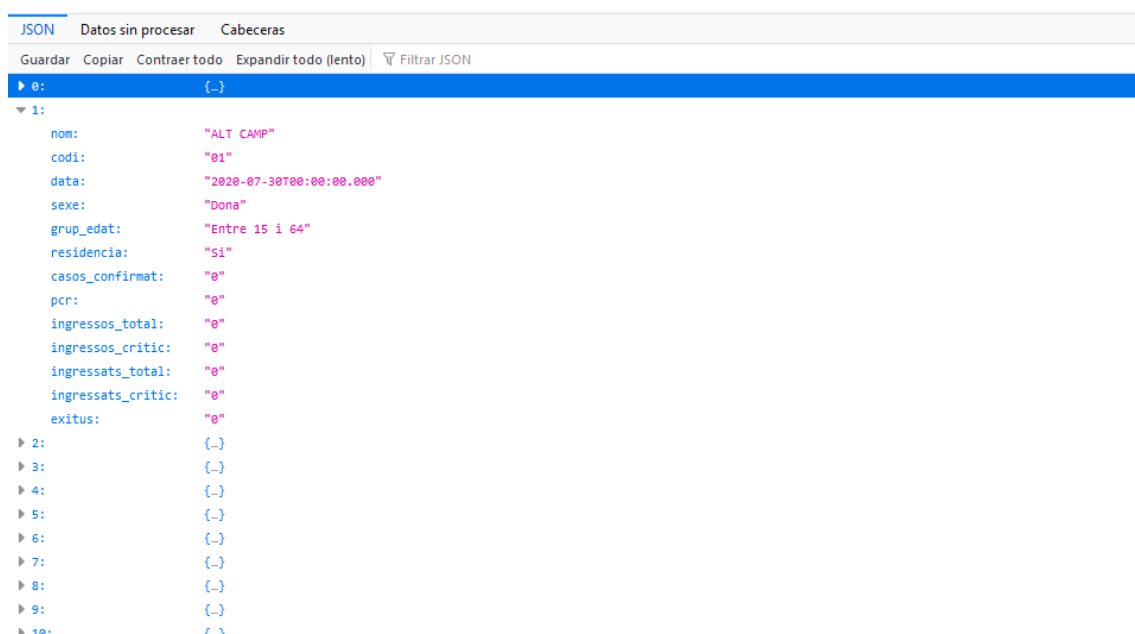
4.1.2 Application on the COVID-19 situation in Alt Empordà

Given the situation of the world pandemic due to COVID-19 and considering the objectives of the world health observatory, the Indika observatory opted for creating a web application with the available updated information about the situation of the Alt Empordà case regarding the pandemic. The objective of the application is to provide information that is relevant to the population and to the territorial agents to improve decision making and the implementation of the

Results

required actions. The application obtains information through the API provided by the Catalonia open data portal (Generalitat de Catalunya, 2020).

The API for obtaining data is a web service that allows the developers to consult the database following the cURL file transfer protocol (Stenberg, 2020). This protocol uses a uniform resources localiser (URL) to indicate what data is required to be accessed. For example, if we access the URL <https://analisi.transparenciaCatalonia.cat/resource/c7sd-zy9j.json> via a browser, we will access a sample of the county's daily COVID-19 data (Departament de Salut, 2020) in JSON format. Figure 18 shows the result of accessing the aforementioned link.



```
JSON  Datos sin procesar  Cabeceras
Guardar Copiar Contraer todo Expandir todo (lento) Filtrar JSON
▶ 0: {}
▼ 1:
  nom: "ALT CAMP"
  codi: "01"
  data: "2020-07-30T00:00:00.000"
  sexe: "Dona"
  grup_edat: "Entre 15 i 64"
  residencia: "si"
  casos_confirmat: "0"
  pcr: "0"
  ingressos_total: "0"
  ingressos_critic: "0"
  ingressats_total: "0"
  ingressats_critic: "0"
  exitus: "0"
▶ 2: {}
▶ 3: {}
▶ 4: {}
▶ 5: {}
▶ 6: {}
▶ 7: {}
▶ 8: {}
▶ 9: {}
▶ 10: {}
```

Figure 18 Screenshot of the daily COVID-19 data by county of the Health Department. Author's own elaboration.

This file transference protocol enables specific consultations to be made by adding URL parameters. In the daily COVID-19 data set for the county (Departament de Salut, 2020) we can specify what county we are interested in, making use of the parameter code. Therefore, if we want to access the Alt Empordà data we would use the URL:

<https://analisi.transparenciaCatalonia.cat/resource/c7sd-zy9j.json?codi=02>

Results

where code = 02 is the identifier of the “Alt Empordà” county. Each data set has some specific parameters. We can consult the options available in the documentation (Generalitat de Catalunya, 2020). Table 7 Daily Covid-19 data by county” shows the fields that make up the Table with daily information about COVID-19 in the Catalan counties from which the income and outgoing calculations are made. The information of this data set is updated with an approximately three-day delay.

Table 7 Daily Covid-19 data by county

Name of the variable	Description	Type
NAME	Name of the county	Text Plan
CODI	Codi de comarca	Text Plan
DATA	Data	Date and time
SEX	Binary field of breakdown of data: includes options MAN and WOMAN	Text Plan
GROUP_AGE	Data breakdown field: includes options: Under 15 years Between 15 and 64 years Between 65 and 74 years Over 75 years	Text Plan
RESIDENCE	Binary field of data breakdown: includes options Yes (population resident in old people’s homes) No (population not resident in old people’s homes)	Text Plan
CASES_CONFIRMED	Numeric field with data of confirmed cases of Covid-19 (PCR+ indicates that	Number

Results

	it is an active case and the date the case started is important for epidemic monitoring. We do not use the COVID-19 cases confirmed by rapid serological tests or by ELISA)	
PCR	Numeric field with data on PCR Tests carried out (The PCR test to detect SARS-CoV-2 can be carried out at different levels, including primary health care, hospital, work, public health, etc. The vast majority are registered in the RSA. From ECAP, the results of the PCR tests carried out on patients in primary private health care are recovered from ECAP. PCRs are generally used to confirm a suspected case of COVID-19, to screen the personal contacts of a confirmed case, or for monitoring purposes. The test results can take between a few hours some days, depending on the place where it was taken. The result of the PCR is imputed on the date the test was carried out, and the figure can increase retrospectively as the laboratory results are received)	Number
ADMISSIONS_TOTAL	Numeric field with the data on the registered admissions (Number of new hospital admissions due to COVID-19 there have been during the period described. These are imputed to the	Number

Results

	territory where the RCA the patient has been assigned to is)	
ADMISSIONS_CRITIC	Numeric field with data of admissions to the registered intensive care units (Number of new patients with COVID-19 admitted to intensive care during the period described. These patients are included in the general admissions, but in the previous days (to be admitted to intensive care, patients must have first been admitted to the hospital). These cases are imputed to the territory where the patient to is assigned to the RCA)	Number
EXITUS	Numeric field with data on the deaths registered (There can be a certain delay in when the deaths are notified. The deaths are imputed on the day of the death and not on the day they are notified. This is why they can increase in number retrospectively)	Number

Source: Department of Health of the Government of Catalonia (2021b)

Information was extracted from the data set “Registry of COVID-19 cases in Catalonia. Segregation by sex and municipality” (Department of Health of the Government of Catalonia, 2021d), which contains up-to-date daily information on the number of cases in each municipality. Only the cases with positive PCRs were used to create the application. Rapid tests, ELISA, and others were not taken into account.

Results

Table 8 Structure of the Table “Registry of COVID-19 cases in Catalonia”

Name of the variable	Description	Type
TipusCasData	Date	Date and time
ComarcaCodi	Region code	Text Plan
ComarcaDescripcio	Region	Text Plan
MunicipiCodi	Municipality code	Text Plan
MunicipiDescripcio	Municipality	Text Plan
SexeCodi	Gender code (0-male, 1-female)	Text Plan
SexeDescripcio	Gender	Text Plan
TipusCasDescripcio	Result	Text Plan
NumCasos	Number of cases	Number

Source: Department of Health of the Government of Catalonia (2021d)

Using the R programming language (R Core Team, 2020), the RSocrata (Devlin *et al.*, 2019) package and the tidyverse package (Wickham *et al.*, 2019), the “Situation of COVID-19 in Alt Empordà” application downloads the information of interest and transforms it into a suitable form for later processing to create the graphics or show the number of confirmed cases.

The user interface (UI) and the server logic were implemented with the Shiny packages (Chang *et al.*, 2020) of R and HTML tools (Cheng *et al.*, 2020). Shiny is a framework for creating web applications that make use of a reactive programming paradigm. The interface was created using the Bootstrap working

Results

framework. Bootstrap is a framework that gives the user interface a specific aesthetic using HTML language in a simple way.

In the middle of the user interface (Figure 19 Application "Situation of COVID-19 in Alt Empordà") there is a map of Alt Empordà, divided by municipalities. On the right, there is a panel that allows the data to be shown on the map to be selected. At the top of the panel, we can select whether we want to see the absolute numbers or rates per 100,000 inhabitants. Just underneath, we can select the time framework of the data. The lower panel shows the figures for the whole of Alt Empordà, Girona, and Catalonia. It also shows two graphics with the historical record of the selected values. The raw data and the accumulated data are at the top and bottom, respectively.

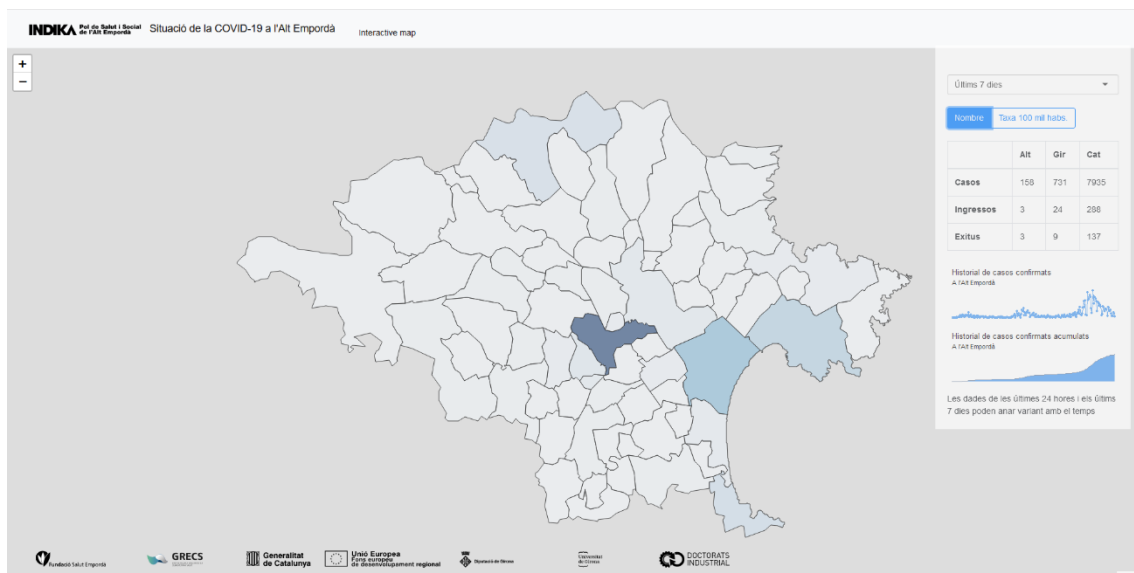


Figure 19 Application "Situation of COVID-19 in Alt Empordà". Source: Author's own elaboration.

The application, which can be hosted on the Shinyapps.io platform in the link <https://fundacio-salut-emporda.shinyapps.io/covid-app/>, is structured in the following way:

- **Data:** The directory that stores the information downloaded from the open data platform.
 - `Altemporda.rds`: Map of the municipalities of Alt Empordà
 - `Catalonia.rds`: Contains the values for the daily confirmed cases, hospital admissions, and deaths in Catalonia for the last 2 weeks.

Results

- Comarca.rds: Contains the historical data for the county of Alt Empordà on the number of positive COVID-19 cases confirmed by a positive PCR.
- Creds.rds: File that stores the credentials of the open data platform to identify who uses it and to ensure that this use is appropriate.
- Girona.rds: Contains the values for the daily confirmed cases, hospital admissions, and deaths in Girona for the last 2 weeks.
- Municipis.rds: Contains information on the number of positive COVID-19 cases confirmed by a positive PCR in Alt Empordà.
- Municipis.cat.rds: Contains the number of confirmed COVID-19 cases for all of Catalonia.
- Municipis-gir.rds: Contains the number of confirmed COVID-19 cases for the province of Girona.
-
- **R:**
 - 00-dades-per-comarca.R: This file is responsible for downloading the information and storing the files in the Data directory.
- **www:**
 - Css: This directory contains a file for improving the aesthetic of the application.
 - Img: This directory contains the images used in the application. In this case, it contains the logotypes of the foundation, the Health and Social Pole of the Alt Empordà, and the university, among others.
 - Js: Contains functions to make the map interactive and to show information when the user clicks on the municipalities.
- **App.R:** The app.R file contains the R code responsible for controlling the server and for showing the information on the screen.

Covid-app.Rproj: A file that defines the directory as an R project.

4.1.3 Key Indicators for Alt Empordà

Key indicators for Alt Empordà (Franquet et al., 2020) is a collection of more than 30 indicators for Alt Empordà, the objective of which is to contextualise the information on the territory; to be the basis from which to make decisions and to

Results

4.1.4 Website of the Indika project

Table 9 shows the data related to the traffic on the Indika project website between January and June. The column “Number of visits” shows the total number of visits to the website, the column “Number of visitors” shows the number of different visitors that have used it, and the column “pages” shows the number of pages visited.

Table 9 Traffic on the Indika website between January and June.

Year	Month	Number of visitors	Number of visits	Pages
2021	January	108	203	2,405
2021	February	188	391	1,074
2021	March	311	679	4,732
2021	April	554	981	2,808
2021	May	563	1,110	2,608
2021	June	363	583	1,144

4.1.5 Other media

The activity of the observatory has been reported several times in the press and on the radio, which can be found in the Appendix named Press releases.

4.2 EVENTR OF R PACKAGE

This section describes how to use the eventr package to implement a data model based on events, the objective of which is to structure an individual’s medical

Results

information, including their date of birth, date of death, and different anthropometric data.

The first step is to define what information needs to be stored and how this will be done. For this example, the following information needs to be registered: date of birth, date of death, and anthropometric data consisting of height, weight, and body mass index. Once the information we need is known, the next step is to define the structure of the individual object.

4.2.1 Structure of the individual

In this example, the individual object is defined as a list of three elements: `birth_date`, `death_date` and `anthropometric_measures`. The `anthropometric_measures` element is a list with three more elements: `weight`, `height` and `body_mass_index`.

```
R > library(eventr)
R > new_individual <- function() {
R +   return(
R +     list(
R +       birth_date = NA,
R +       death_date = NA,
R +       anthropometric_measures = list(
R +         weight = NA,
R +         height = NA,
R +         body_mass_index = NA
R +       )
R +     )
R +   )
R + }
R >
R > individual <- new_individual()
```

4.2.2 Create the specifications.

Given the proposed objectives, as a minimum requirement the specifications of three events must be created: the event Birth, the event Death, and the event

Results

Anthropometric Measures. The specification of an event is a function that returns an event type object.

```
R > specification <- function(<params>) {  
R +   return(  
R +     event(  
R +       id = <event_id>,  
R +       type = <event_type>,  
R +       time = <event_time>,  
R +       <event_params>  
R +     )  
R +   )  
R + }
```

The parameters of the specification of the event `add_birth_spec` are `event_id` and `birth_date`. The parameter `event_id` is a character that allows a specific event to be identified. It must be a single value. Should an identifier not be provided, the function `event()` assigns a unique universal identifier (`uuid`) using the `uuid` packet. The parameter `birth_date` must be a date in the format `YYYY-MM-DD`. For this example, the parameter `time` will be the same as the date of birth `birth_date`.

```
R > add_birth_spec <- function(event_id, birth_date) {  
R +   return(  
R +     event(  
R +       id = event_id,  
R +       type = 'BIRTH',  
R +       time = birth_date,  
R +       birth_date = birth_date  
R +     )  
R +   )  
R + }
```

The parameters of the specification of the event `add_death_spec` are `event_id` and `death_date`. The parameter `event_id` is the same as the specification `add_birth_spec`, and the parameter `death_date` is similar to the parameter `birth_date` of this specification.

Results

```
R> add_death_spec <- function(event_id, death_date){
R+   return(
R+     event(
R+       id = event_id,
R+       type = 'DEATH',
R+       time = death_date,
R+       death_date = death_date
R+     )
R+   )
R+ }
```

The specification of the event `add_anthropometric_measures_spec` has three parameters: `event_id` (the same as the previous ones), `weight` and `height`. The parameter `weight` represents the weight of the individual in kilograms. The parameter `height` represents the height of an individual in meters. Given that the body mass index is calculated using these two values, it does not need to be included in the specification of the event. The `handler` will be responsible for doing the calculations.

```
R> add_anthropometric_measures_spec <-
R+ function(event_id, time, weight, height){
R+   return(
R+     event(
R+       id = event_id,
R+       time = time,
R+       type = 'ADD_ANTHROPOMETRIC_MEASURES',
R+       weight = weight,
R+       height = height R+   )
R+   )
R+ }
```

4.2.3 Creating the handlers

The function of the handlers is to transform the attributes of an object (in this case, the object `individual`) from the information contained in an event. There must be a handler (`handler`) for each type of event created.

Results

The handler of the event Birth (`birth_handler`) is responsible for modifying the attribute `birth_date` of the individual object and assigning it the value `birth_date` of the input event.

```
R> birth_handler <- handler(  
R+   type = 'BIRTH',  
R+   FUN = function(obj, event){  
R+  
R+     obj$birth_date <- get_body_attr(event, 'birth_date')  
R+     return(obj)  
R+  
R+   })
```

The handler of the event DEATH (`death_handler`) is responsible for modifying the attribute `death_date` of the individual object and assigning it the value `death_date` of the input event.

```
R> death_handler <- handler(  
R+   type = 'DEATH',  
R+   FUN = function(obj, event){  
R+  
R+     obj$death_date <-  
R+     get_body_attr(event, 'death_date')  
R+     return(obj)  
R+  
R+   })
```

The handler of the event `ADD_ANTHROPOMETRIC_MEASURES` (`anthropometric_measures_handler`) is somewhat more complex. This handler assigns the values `weight` and `height` of the event to the `weight` and

Results

height attributes of the individual object, based on which the value of the attribute `body_mass_index` is calculated,

```
R> anthropometric_measures_handler <- handler(  
R+   type = 'ADD_ANTHROPOMETRIC_MEASURES',  
R+   FUN = function(obj, event){  
R+  
R+     weight = get_body_attr(event, 'weight')  
R+     height = get_body_attr(event, 'height')  
R+     body_mass_index = weight / height^2  
R+  
R+     obj$anthropometric_measures$weight = weight  
R+     obj$anthropometric_measures$height = height  
R+     obj$anthropometric_measures$body_mass_index=  
body_mass_index  
R+  
R+     return(obj)  
R+  
R+   }  
R+ )
```

4.2.4 Create the dispatcher,

The function `dispatcher()` creates a `dispatcher` type object based on the handlers (`handlers`). This object has got two elements: the list of `handlers` and the function `dispatch()`, which is the function responsible for sending each event to its respective handler.

```
R> the_dispatcher <- dispatcher(  
R+   handlers = handlers_list(  
R+     birth_handler,  
R+     death_handler,  
R+     anthropometric_measures_handler  
R+   )  
R+ )  
R>  
R> str(repartidor, max.level = 1)
```

Results

```
List of 2
 $ handlers:List of 3
  ..- attr(*,"class")= chr [1:2] "handlers_list" "list"
 $dispatch:function(obj = NULL, events, accumulate = FALSE)
  - attr(*, "class")= chr [1:2] "dispatcher" "list"
```

4.2.5 Creating the occurrences of the events

The occurrences of the events are the events that have taken place and have changed the state of the individual object. Four occurrences have been created for this example. The first corresponds to the Birth of the individual, the second and third to the times when the anthropometric measures were taken, and last to the death of the individual.

```
R> the_events <-
R+   add_birth_spec(
R+   'first-event',
R+   birth_date = as.Date('1936-11-09')
R+ ) +
R+   add_anthropometric_measures_spec(
R+   'second-id',
R+   as.Date('1951-01-26'),
R+   65,
R+   1.72
R+ ) +
R+   add_anthropometric_measures_spec(
R+   'third-id',
R+   as.Date('1982-05-03'),
R+   70,
R+   1.72
R+ ) +
R+   add_death_spec(
R+   'fourth-id',
R+   death_date = as.Date('2018-05-11')
R+ )
```

Results

4.2.6 Executing the events

Once the events are created, we can calculate the state of the individual object using the function `dispatch()` of the object dispatcher.

```
R> dispatch <- get_dispatch(the_dispatcher)
R> individual_01 <- dispatch(obj = individual, events = the_events)
R> individual_01

$birth_date
[1] "1936-11-09"

$death_date
[1] "2018-05-11"

$anthropometric_measures

$anthropometric_measures$weight
[1] 70

$anthropometric_measures$height
[1] 1.72

$anthropometric_measures$body_mass_index
[1] 23.66144
```

It may be of interest to obtain the state of the individual in each of the events, which is why the function `dispatch()` allows this to be indicated using the parameter `accumulate`. The result is a list with as many elements as events, plus one.

```
library(dplyr)
library(purrr)

individual_02 <- dispatch(obj = individual, events = the_events,
  accumulate = TRUE)

individual_to_df <- function(individual){
```

Results

```
df1 <- data.frame(  
  birth_date = individual$birth_date,  
  death_date = individual$death_date  
)  
  
df2<- as.data.frame(individual$anthropometric_measures)  
  
return(cbind(df1, df2))  
  
}  
  
individual_02 %>%  
  purrr::map(individual_to_df) %>%  
  bind_rows()  
  
birth_date death_date weight height body_mass_index  
1 <NA> <NA> NA NA NA  
2 1936-11-09 <NA> NA NA NA  
3 1936-11-09 <NA> 65 1.72 21.97134  
4 1936-11-09 <NA> 70 1.72 23.66144  
5 1936-11-09 2018-05-11 70 1.72 23.66144
```

4.2.7 Advantages of the use of EDA

This section describes how to use the functions and structures implemented in `eventr` to make the most of the advantages of an EDA.

4.2.7.1 Obtaining the auditory registry

In the implementation carried out, the auditory registry is a list of events (`event_list`) that stores all the events that have taken place concerning an object of interest (`the_events` in the previous example).

By defect, an `event_list` is structured as a list. It may be of interest to transform this structure into a `data.frame`. This is why it is recommended to transform the `headers` and the (`bodies`) separately. Given that the `bodies` will depend on the structure of the events, it is recommended to create an auxiliary function (`aux_foo`) to transform these codes into a `data.frame`, as shown below:

Results

```
# Headers to data.frame
the_headers <- get_header(the_events) %>%
  lapply(as.data.frame) %>%
  bind_rows()

# Auxiliar function
aux_foo = function(x) as.data.frame(x)

# bodies to data.frame
the_bodies <- get_body(the_events) %>%
  lapply(aux_foo) %>%
  bind_rows()

audit_trail <- bind_cols(the_headers, the_bodies)
audit_trail
```

	id	type	time
1	first-event	BIRTH	1936-11-09
2	second-id	ADD_ANTHROPOMETRIC_MEASURES	1951-01-26
3	third-id	ADD_ANTHROPOMETRIC_MEASURES	1982-05-03
4	fourth-id	DEATH	2018-05-11

	birth_date	weight	height	death_date
1	1936-11-09	NA	NA	<NA>
2	<NA>	65	1.72	<NA>
3	<NA>	70	1.72	<NA>
4	<NA>	NA	NA	2018-05-11

4.2.7.2 Testing the events

Let's assume that we want to evaluate the correct functioning of the handler `anthropometric_measures_handler`. This is why we will create a new individual with the values that their attributes must take if the handler for evaluating functions correctly.

```
R> expected_individual <- new_individual()
```

Results

```
R> expected_individual$anthropometric_measures$weight <- 65
R> expected_individual$anthropometric_measures$height <- 1.72
R> expected_individual$anthropometric_measures$body_mass_index <- 65
/ 1.72^2
```

Later, an occurrence with the test values is created (`test_ocurr`).

```
R> test_ocurr <- add_anthropometric_measures_spec(
R+   event_id = 'second-id',
R+   time = as.Date('1951-01-26'),
R+   weight = 65,
R+   height = 1.72
R+ )
```

Using the function `get_fun()`, we can extract the function of the handler, which by defect must have two parameters `obj` and `event`.

```
R> foo <- get_fun(anthropometric_measures_handler)
R>
R> test_individual <- foo(new_individual(), test_ocurr)
```

If everything functions as planned, the objects `expected_individual` and `test_individual` must be identical. As a result, the function `identical()` should return the value `TRUE`.

```
R> identical(test_individual, expected_individual)
[1] TRUE
```


Results

4.2.7.3 *Retroactive events*

The retroactive events are events that have an effect on a previous temporal moment. This property is especially useful to recalculate the state of an object when there has been some type of error in inputting the information. To this effect, there is no need to worry about the collateral effects that this will have had on the final state. Let's suppose that a user inputs the height of an individual in centimetres rather than in metres (which is the defect unit).

```
correct_birth_date <- add_birth_spec(  
  event_id = 'first-event',  
  birth_date = as.Date('1936-11-09')  
)  
  
incorrect_event <- add_anthropometric_measures_spec(  
  event_id = 'second-id',  
  time = as.Date('1951-01-26'),  
  weight = 65,  
  height = 172 # Incorrect measurement  
)  
  
incorrect_events_list <- event_list(correct_birth_date,  
incorrect_event)  
  
incorrect_individual <-  
  repartir(  
    obj = new_individual(),  
    events = incorrect_events_list,  
    accumulate = FALSE  
  )  
  
individual_to_df(incorrect_individual)  
  
  birth_date  death_date  weight  height  body_mass_index  
1 1936-11-09  NA         65     172     0.002197134
```

Results

As can be seen in the previous results, the attribute `body_mass_index` has not been calculated correctly. By changing the incorrect event for the event with the correct values, the state of the individual can again be calculated.

```
correct_event <- add_anthropometric_measures_spec(
  event_id = 'second-id',
  time = as.Date('1951-01-26'),
  weight = 65,
  height = 1.72 # Correcte measurement
)

correct_events_list <- incorrect_events_list
correct_events_list[[2]] <- correct_event

correct_individual <-
  repartir(
    obj = new_individual(),
    events = correct_events_list,
    accumulate = FALSE
  )

individual_to_df(correct_individual)

  birth_date death_date weight height body_mass_index
1 1936-11-09  NA         65     1.72     21.97134
```

4.2.7.4 *Alternative states*

The alternate states allow the state of an individual in different temporal moments or the state of an individual using two different dispatchers to be obtained. In this case, what is shown is how to obtain an individual in a previous temporal moment, or on 27 January 1951, based on the list of events of the object `the_events`.

```
the_events_time <-
  lapply(the_events, get_time) %>%
  unlist()
```

Results

```
alternative_events_list <-  
  the_events[the_events_time <= as.Date('1951-01-27')]  
  
alter_individual <- repartir(  
  obj = new_individual(),  
  events = alternative_events_list,  
  accumulate = FALSE  
  ) %>%  
  individual_to_df()  
  
alter_individual  
  
  birth_date  death_date  weight  height  body_mass_index  
1 1936-11-09      NA         65     1.72     21.97134
```

4.2.8 Application to a real case

This section describes how we can use the proposed architecture to structure medical information from the electronic health registry of Catalonia (Spain). The data contain information about Alt Empordà, a territorial unit located in the north of Catalonia on the border with France, and were obtained from the Programme for Data Analysis for Research and Innovation in Health (PADRIS), from the Catalan Government's Quality and Health Evaluation Agency of Catalonia (AQuAS).

The data set used contains information about people that suffer from type 2 diabetes mellitus (DM2) and, where applicable, the date of diagnosis. Furthermore, it contains information like the date of birth, if the person is alive or not, the date of death (where applicable), and the sex of the individual. There are 66,292 individuals in the data set, of which 32,802 are men and 33,490 are women. The period of study ran from 1 January 2008 to 31 December 2018. Of the total set of individual, 4,395 had DM2 during the study period, at the end of which 3,391 individuals had died. The data are available in the R packet, which we can access via the data function().

Results

```
data (Alt_Empordà)
head (Alt_Empordà)

  id sex dm2 dm2_date birth_year is_dead death_date
1 26000153 0 1 2002.12.05 1945 0 <NA>
2 27288814 0 1 2010-04-07 1936 0 <NA>
3 27161071 1 1 2010-08-09 1934 0 <NA>
4 25517840 0 1 2010-03-16 1938 1 2019.02.01
5 22650869 1 1 2010-11-19 1943 0 <NA>
6 27455596 1 1 2013-01-18 1938 0 <NA>
```

Shown below is how to transform the data from the previous table into a structure directed by events.

4.2.8.1 Structure of the individual

For this example, an individual is defined as a list of four attributes: an identifier (`id`), the sex variable (`sex`) which can be either the value 0 (man) or the value 1 (woman), `is_dead` indicates whether the individual is dead (1) or not (0) and `dm2` if the individual in question has type 2 diabetes mellitus or not.

```
library(eventr)

new_individual <- function() {

  return(
    list(
      birth_date = NA,
      death_date = NA,
      anthropometric_measures = list(
        weight = NA,
        height = NA,
        body_mass_index = NA
      )
    )
  )
}
```

Results

```
individual <- new_individual()
```

4.2.8.2 Create the specifications of the events.

Three specifications are defined: Birth (`birth_spec`), Death (`death_spec`) and type 2 diabetes mellitus (`dm2_spec`).

The specification of the event Birth (`birth_spec`) has two input parameters. The parameter `birth_year` indicates the year when the individual was born. The parameter `sex` indicates the sex of the individual and can take one of two values: man (0) or woman (1).

```
birth_spec <- function(birth_year, sex ){
  return(
    event(
      id = uuid::UUIDgenerate(),
      type = 'BIRTH',
      time = as.character(birth_year),
      is_dead = FALSE,
      sex = sex
    )
  )
}
```

The specification of the event Death (`death_spec`) generates an identifier of the event using the function `UUIDgenerate()`, assigns the type of event `DEATH`, and the time of the event such as the date of death `death_date`. The value `TRUE` is assigned to the parameter `defect is_dead` when this specification is used.

```
death_spec <- function(death_date){
  return(
```

Results

```
event(  
  id = uuid::UUIDgenerate(),  
  type = 'DEATH',  
  time = as.character(death_date),  
  is_dead = TRUE  
)  
)  
}
```

The specification of type 2 assigns the type `DIABETS_MELLITUS_TYPE_2` to the event and assigns the value of the attribute `dm2` to `TRUE` by defect.

```
dm2_spec <- function(event_time) {  
  
  return(  
    event(  
      id = uuid::UUIDgenerate(),  
      type = 'DIABETES_MELLITUS_TYPE_2',  
      time = as.character(event_time),  
      dm2 = TRUE  
    )  
  )  
}
```

Once the specifications are defined, the handlers are implemented.

4.2.8.3 *Create the handlers.*

The handlers are responsible for transforming the parameters of an event into an individual's attributes.

The handler `birth_handler` is responsible for manipulating the `BIRTH` type events. The attributes `is_dead` and `sex` are assigned based on the `BIRTH` type events.

Results

```
birth_handler <- handler(  
  type = 'BIRTH',  
  FUN = function(obj, event){  
    obj$is_dead <- get_body_attr(event, 'is_dead')[[1]]  
    obj$sex <- get_body_attr(event, 'sex')[[1]]  
    return(obj)  
  })
```

The handler `death_handler` is responsible for manipulating the `DEATH` type events. It modifies the attribute `is_dead` of the individual.

```
death_handler <- handler(  
  type = 'DEATH',  
  FUN = function(obj, event){  
    obj$is_dead <- get_body_attr(event, 'is_dead')[[1]]  
    return(obj)  
  })
```

The handler `dm2_handler` is responsible for manipulating the `DIABETES_MELLITUS_TYPE_2` type events. It modifies the attribute `dm2` of the individual.

```
dm2_handler <- handler(  
  type = 'DIABETES_MELLITUS_TYPE_2',  
  FUN = function(obj, event){  
    obj$dm2 <- get_body_attr(event, 'dm2')  
    return(obj)  
  })
```

Results

4.2.8.4 *Creating the dispatcher*

Once the handlers are created, the new dispatcher is created using the function `dispatcher()`.

```
the_dispatcher <- dispatcher(  
  handlers = handlers_list(  
    birth_handler,  
    death_handler,  
    dm2_handler  
  )  
)
```

4.2.8.5 *Creating the occurrences of the individuals*

As shown below, a list is created with the events for each individual. Each element of the individuals list represents an individual from the original data set and is made up of a list of events (`event_list`).

```
> # split data.frame by individual  
> individuals <- split(alt_emporda, f = alt_emporda$id)  
>  
> head(individuals, 3)  
$`12602285`  
      id sex dm2 dm2_date birth_year is_dead death_date  
24002 12602285 1 0 <NA>      1971      0 <NA>  
  
$`12602295`  
      id sex dm2 dm2_date birth_year is_dead death_date  
24003 12602295 0 0 <NA>      1992      0 <NA>  
  
$`12622895`  
      id sex dm2 dm2_date birth_year is_dead death_date  
24004 12622895 0 0 <NA>      2018      0 <NA>
```


Results

The events for each of the elements of the individuals list are created.

```
individuals <- lapply(individuals, function(x){

# birth event

the_events <- birth_spec(
  birth_year = unique(x$birth_year),
  sex = ifelse(
    unique(x$sex) == 1,
    'woman', ifelse(
      unique(x$sex) == 0,
      'man',
      'unknown'
    )
  )
)

# create death event if is_dead == 1

if(x$is_dead == 1){

  the_events <- the_events +
  death_spec(death_date = unique(x$death_date)[[1]])

}

# create diabetes mellitus type 2 event if dm2 == 1

if(x$dm2 == 1){

  the_events <- the_events +
  dm2_spec(
    event_time = unique(x$dm2_date)
  )

}

if(class(the_events)[1] == "event") the_events <-
event_list(the_events)

return(the_events)

}
```

Results

```
})
```

Once the individual's events are created, we can use the eventr functions.

4.2.8.6 Execute the events for each individual.

```
repartir <- get_dispatch(the_dispatcher)
ids <- names(individuals)

li <- list()

for(id in ids){

  x <- individuals[id]
  li[[id]] <- repartir(
    obj = new_individual(id = names(x)),
    events = x[[1]],
    accumulate = FALSE
  ) %>% as.data.frame()

}

li %>% bind_rows() %>% head(6)
```

	id	sex	is_dead	dm2
1	12602285	woman	FALSE	FALSE
2	12602295	man	FALSE	FALSE
3	12622895	man	FALSE	FALSE
4	12622896	woman	FALSE	FALSE
5	12622899	woman	FALSE	FALSE
6	12623072	woman	FALSE	FALSE

Results

In the previous example, the value `accumulate = FALSE` was used, and so the final result is a simple row per individual. If we are interested in obtaining a row for each event that has taken place, we can use the parameter `accumulate = TRUE`, as shown below.

```
second_list <- list()

for(id in ids){

  x <- individuals[id]
  second_list[[id]] <- repartir(
    obj = new_individual(id = names(x)),
    events = x[[1]],
    accumulate = TRUE
  ) %>%
  purrr::map(as.data.frame) %>%
  bind_rows() %>%
  # Elim first row which contains default individual
  .[-1,]

}

second_list <- second_list %>% bind_rows()
second_list[25:31,]

      id      sex  is_dead  dm2
25 12752496  man    FALSE  FALSE
26 12752497  man    FALSE  FALSE
27 12752498 woman  FALSE  FALSE
28 12752576  man    FALSE  FALSE
29 12752583  man    FALSE  FALSE
30 12752583  man     TRUE  FALSE
31 12752638  man    FALSE  FALSE
```

The user with identifier 12,752,583 has two rows given that in the end the study was dead.

Results

4.2.8.7 Obtain the Audit Trail

The auditory registry allows us to revise all the events that have taken place for an individual and which attributes have been modified in each of the events. This is why it is a useful tool to identify code failures. Shown below is the way to obtain the auditory registry for the set of individual data.

```
audit_trail <- lapply(individuals, function(occurrences){

  if(length(occurrences) == 1){

    headers <-
      get_header(occurrences) %>%
      as.data.frame() %>%
      rename(event_id = id)
    bodies <-
      get_body(occurrences) %>%
      as.data.frame()

    return(bind_cols(headers, bodies))

  } else {

    if(length(occurrences) > 1){

      headers <- lapply(occurrences, function(x) {
        get_header(x) %>%
          as.data.frame() %>%
          rename(event_id = id)
      }) %>% bind_rows()

      bodies <- lapply(
        X = occurrences,
        FUN = function(x) {
          get_body(x) %>% as.data.frame()
        }) %>% bind_rows()

      return(bind_cols(headers, bodies))

    }

  }

}
```

Results

```
}  
  
audit_trail %>%  
  bind_rows(.id = 'individual_id') %>%  
  head()  
  
  individual_id      event_id type time is_dead sex dm2  
1      12602285 94624a1b-4dad-4366-9a3e-804c7fc82b4e BIRTH 1971 FALSE  
woman NA  
2      12602295 6dbdd9aa-dfcb-4a77-bfe3-7fbbb7bf206f BIRTH 1992 FALSE  
man NA  
3      12622895 2bbc68f3-b055-4174-8f62-16ffc8cc4f10 BIRTH 2018 FALSE  
man NA  
4      12622896 2266137f-703c-4d02-ab13-166543ee71f7 BIRTH 2018 FALSE  
woman NA  
5      12622899 48828e85-d8eb-46ec-bca6-09bc5d90e627 BIRTH 2018 FALSE  
woman NA  
6      12623072 8b493907-e2de-47e8-b02f-91dce32f4fe8 BIRTH 2009 FALSE  
woman NA
```

As shown in the previous code, the auditory trail provided by the individual shows the identifier of the event (`event_id`), the type of event that has taken place (`type`), the time when it took place (`time`), and the values of the attributes in each of the events. For the example shown, the modified attributes are `is_dead` and `sex`, given that the events shown as a result of the function `head` belong to the `BIRTH` type events.

The number of downloads from CRAN (Institute for Statistics and Mathematics of WU (Wirtschaftsuniversität Wien), 2021) was used to evaluate the impact of the **eventr** package. The package `cranlogs` (Csárdi, 2021) was used to obtain the number of downloads. In the period between 26 May 2020 (when the CRAN repository was uploaded) and the time of writing, the `eventr` package was downloaded 3,415 times. This means that on average it was downloaded 293 times a month or 11 times a day. Table 10 shows the number of monthly downloads.

Results

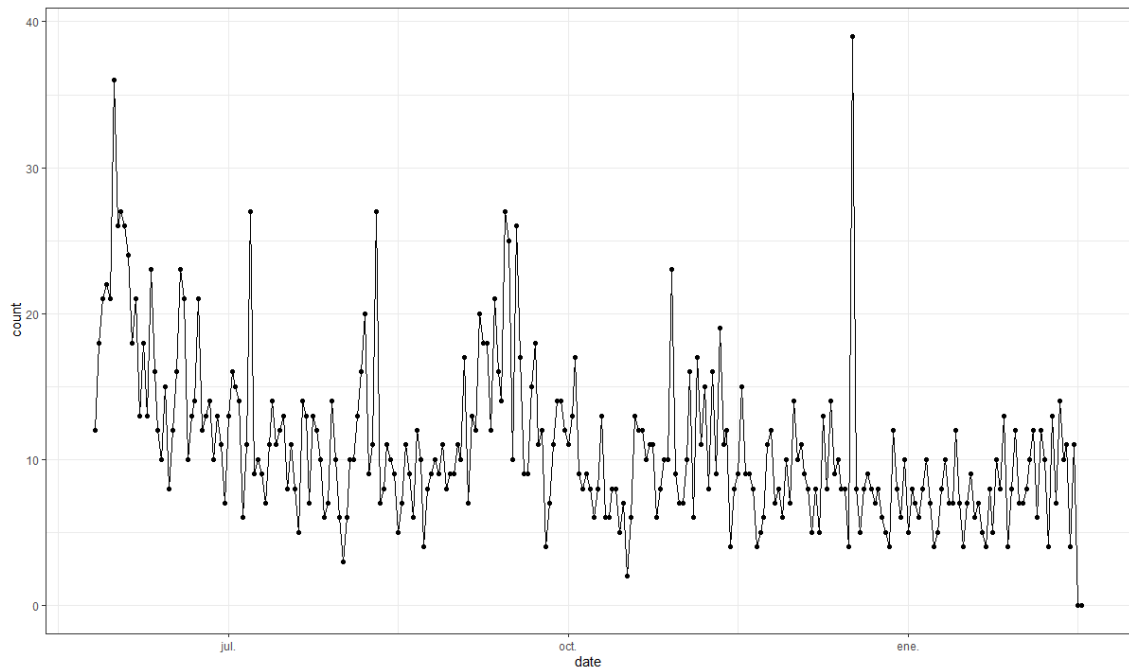


Figure 21 Downloads of the eventr package from 2020-05-26 to 2021-06-04. Source: Author's own elaboration based on data obtained through the cranlogs package.

Table 10 Number of monthly downloads of the eventr package from the CRAN repository. Source: Author's own elaboration based on data obtained through the cranlogs package.

Year	Month	N
2020	May	130
2020	June	480
2020	July	342
2020	August	307
2020	September	430
2020	October	292
2020	November	295

Results

2020	December	287
2021	January	228
2021	February	293
2021	March	322

Source: Author's own elaboration.

4.3 INDIKA DATA REPOSITORY

The Indika data repository hosted on the GitHub website can be found in the link <https://github.com/alvarofranq/indika-data-repository>. Figure 22 shows the main page of the repository. The repository has 2 indicators at the moment of writing the thesis.

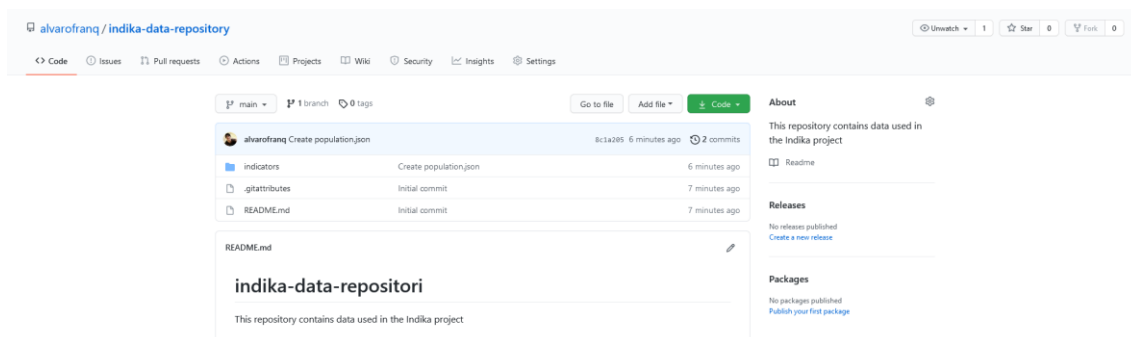


Figure 22 Home page of the Indika data repository. Source: Author's own elaboration.

The information on the indicators can be accessed using the R commands and the package *jsonlite*, as shown below:

```
> URL <- 'https://raw.githubusercontent.com/alvarofranq/indika-data-repository/main/indicators/population.json'

> indicator <- jsonlite::read_json(URL, simplifyVector = TRUE)

> head(indicator$data)
  municipalityCode municipalityName mens womens total
1         170010         Agullana  431    432   863
2         170031         Albanyà   79     76   155
```

Results

3	170118	Armentera, l'	478	473	951
4	170123	Avinyonet de Puigventós	818	823	1641
5	170160	Bàscara	521	480	1001
6	172348	Biure	125	116	241

In the previous example, the information corresponding to the indicator *population* containing the number of men and women in each municipality in Alt Empordà can be downloaded.

Similarly, we can use the lists of indicators to download multiple indicators. Below is an example of how to use R to download the indicators included in the list of indicators, in this case the list *peopleAndPopulation*.

```
library(jsonlite)
library(tidyverse)
library(rvest)

> URL <- 'https://raw.githubusercontent.com/alvarofranq/indika-data-
repository/main/indicators-list/peopleAndPopulation.json'
indicators_list <- jsonlite::read_json(URL, simplifyVector = TRUE)

> people_and_population <- lapply(indicators_list,
jsonlite::read_json, simplifyVector = TRUE)

> str(people_and_population, max.level = 2)
List of 2
 $ :List of 2
  ..$ meta:List of 6
  ..$ data:'data.frame': 68 obs. of 5 variables:
 $ :List of 2
  ..$ meta:List of 6
  ..$ data:'data.frame': 126 obs. of 5 variables:
```


Results

5 DISCUSSION

The aim of this thesis was to develop a health and social observatory for Alt Empordà from a scientific perspective, and to provide it with the tools needed for its initial and future function. The following sections are a global discussion of its main results, together with its limitation, strengths, and future research directions. What follows is a start to the discussion process based on three main points: first, the construction of the Indika observatory; second, the construction of the public information repository; and last, the strengths and weaknesses of using an architecture based on events to structure the information, and its implementation in the *eventr* package.

5.1 INDIKA HEALTH AND SOCIAL POLE OF THE ALT EMPORDÀ

Based on the work of Caiaffa *et al.* (2014) as a reference framework, we can evaluate the characteristics of the observatory and see whether they adjust to the concept of conservatory created. Below is a point-by-point description of the appraisal of the work done. The evaluation can be structured in four parts: the mission of the observatory, its governance, the integration of the territorial information, and the evidence generated.

Indika's mission is clear and is aligned with the objectives of sustainable development. It has the capacity to generate knowledge about topics of interest to the county and to facilitate training in the different environments. It is a multidisciplinary observatory, which brings different visions to the data processing and the generation of knowledge. Furthermore, it is capable of producing material on which to base decisions, such as the "Key Indicators for Alt Empordà" (Franquet *et al.*, 2020).

Governance, understood as a way of organising the observatory and the way in which it relates to other entities, is not clear, and neither are the responsibilities of its different members. Who is responsible for each action and the decisions made, and which entities are involved and how (contributing economic and human capital), must be more clearly specified. This is the case of two of the entities involved, Figueres City Council and the county council. It is not clear what

Discussion

contributions each makes are not clear or what individual responsibilities they have over the information published. Another important point regarding governance is managing to diversify how capital is obtained so that not everything depends on a single entity, with the aim of ensuring Indika's long-term sustainability. At the time of delivery of this thesis, Indika had obtained public funding to be able to continue with its activity within the framework of the PECT Innovació Social-EMISSE project.

Indika has a good business intelligence capacity, enabling it to identify and make use of public and private interest resources. The observatory has an information repository that is replicable for other counties and expandible, which facilitates the traceability of the analyses made and enables its integration with other projects that are being carried out at the county level, including the Dipsalut Observatory of Health and Social Inequalities. The reproducibility of the information published is also guaranteed, given that all the analyses were carried out using the R programming language (R Core Team, 2020), which is open and accessible to everyone. The different codes used can be requested directly from the observatory. With the COVID-19 application, Indika has shown that it is able to monitor the state of the territory effectively, and the different press appearances are proof that it has the capability to disseminate the actions carried out.

A flexible working methodology (Fowler i Highsmith, 2001) was used in the process of developing the observatory. The well-used flexible working methodologies used in the development of the software place the focus on the iterative design of the product where the requisites and solutions evolve over time in line with project needs. The purpose of the working teams, which must be self-organised and multidisciplinary, as is the case of Indika, is to resolve the different problems that emerge during the process collaboratively and in a self-organised manner.

In the case of Indika, this process consisted of regular meetings with the observatory team where the short-term objectives to be achieved were decided. These objectives included the creation of the webpage and the different

Discussion

infographics, prioritisation of the actions to be developed, assignation of the tasks among the different participants, and decision making as to their deadlines. These schedulings of actions are known as *sprints*.

The meetings were structured in three main parts: the amalgamation of the results obtained, exposure of the status of the actions developed during that period and, should a problem arise, discussion as to the actions to resolve it; decision making as to the priorities; and last, deciding on the actions for the new *sprint*.

This process also includes carrying out meetings to retrospectively evaluate what has worked well, what has to be approved, and what has not worked, with the aim of improving the weak points and promoting the strong points of the team.

The pandemic has forced the observatory to change its IT priorities. Two main actions have been carried out to this effect: the creation of a web application to be able to see the status of the county, and an infographic summary a year after the first case of COVID-19 confirmed in the county. From these we obtained the learnings described below.

The creation of a specific web application, as is the case of the application mentioned in section 4.1.2, has a resources cost that is too high for the real use made of it. While it was a useful application at the beginning of the pandemic, it was not as useful at the end when there were many other options that offered more detailed information. Given that the data changes as new findings are made and current needs change, a person solely responsible for this task is required to be able to offer an up-to-date application that adds value. The profitability of this type of action must be evaluated considering the many factors involved in an observatory like Indika.

Conversely, the creation of a report in the form of an annual summary as described in section 4.1.1.3 was a very good decision, given that its high media impact enabled us to raise the profile of the observatory and awareness of the situation in the county.

Discussion

Notably, the working methodology used was easily adaptable to the new needs that emerged as a result of the pandemic.

5.2 INDIKA DATA REPOSITORY

The Indika data repository contains relevant information on the territory in JSON format, which can be downloaded from the GitHub (2020) website or directly using the URLs for the different indicators, as shown in section 4.3. The use of the URLs facilitates the exploitation of the data using programming languages such as R (R Core Team, 2020). Thanks to the lists of indicators, multiple indicators can be downloaded simultaneously.

Using the GitHub (2020) platform allows the users to access the information and to actively participate in its update and collection. Furthermore, there is the option of downloading all the information for use at the local level. Any entity or person can add new indicators, ensuring that the structure follows the characteristics described in section 3.2.

This point highlights the need to create information verification mechanisms, which can be the review of the data should they be public domain data, or the need to contact Indika first should the data being processed come from private persons or entities for verification of their authenticity.

Given the characteristics of the repository, who or what entity has made each update can be known in order to give merit where due. Furthermore, given that this repository is open and available to anyone, we can say that the Indika data repository meets the requirements defined in section 3.2.

5.3 EVENTR

The *eventr* package has been published in the *Comprehensive R Archive Network (CRAN)* (Institute for Statistics and Mathematics of WU (Wirtschaftsuniversität Wien), 2021), the reference repository for all the R packages. It must undergo a review process conducted by an expert in the subject to be published in CRAN, which would signify that it is of a high enough quality to be published. Its publication in CRAN would imply that the *eventr*

Discussion

structure meets the requirements of the development community, conferring its validity. Therefore, the documentation of the eventr package is sufficient to know what its use is and how to use it.

The package has also been published on the GitHub open code website and available in this link: <https://github.com/alvarofrang/eventr>. This website allows possible problems derived from the use of eventr to be notified and its functionalities to be added or modified.

The process of creating eventr was structured in three steps: needs identification, design of the implementation, and implementation in R. Identification of the requirements was carried out correctly, focusing the problem on the need to decide how to structure the information to be able to re-use it, making use of different statistical analysis tools.

Regarding the design, the focus was on an architecture based on events, enabling multiple updates and thereby increasing its complexity, or constructing other packages on top, adding to the wealth of the data and their possibilities for use.

As mentioned above, the implementation carried out meets all the requirements defined by the developers' community, as does the required documentation needed for its use, as illustrated by the examples with real data.

One of the problems of the implementation of the package is its high complexity derived from the technical difficulty of implementing architectures based on events in real environments. The transformation of structured data using other architectures can be complex, disincentivising the use of this type of methodology.

6 CONCLUSIONS AND FUTURE WORK

Based on the results obtained in this thesis, we can conclude that:

- The territorial health observatories, such as the ones as the county level, are a good tool to provide the territory with the business intelligence in health needed to take decision. However, collaboration at a higher level is needed to ensure that efforts are not duplicated, and to share technologies and analysis tools with other existing elements. Aside from the function of a county health observatory, efforts must especially focus on creating effective collaboration networks that also generate scientific knowledge on the actions carried out in the territory thanks to the analysis of the available data.
- An architecture based on events facilitates scientific statistical analysis of the information available in the digital medical registries, among others. Its implementation is costly in terms of time and human capital. These two factors need to be considered when using it and developing the processes.
- The creation of an open data repository like Indika must go hand in hand with a process of dissemination that includes the creation of teaching material to educate possible users on how to use and collaborate with the repository. This material must bear in mind the different existing professional profiles, including people with the technical capacity to collaborate directly and the entities that have data but are not able to exploit it adequately, accompanying them in the process of opening information.

Based on the discussion featured above, some points or lines of work, and future lines of research, have been able to be identified, such as:

Conclusions and future work

- Design a communication strategy that uses social networks to disseminate the work done by Indika, especially if the youngest sector of the population of the territory are to be reached.
- To better specify the implication of the different territorial alliances, framing this implication in specific actions that have a direct impact on Alt Empordà.
- To design a collaboration strategy with the different territorial entities. In other words, to redefine and specify how we can work together with other territorial agents, and what we consider to be effective collaboration.
- To advance in a scalable cooperation and work with other territorial or higher-level observatories.
- To design a long-term strategy wherein the most important problems of the territory are identified and that clearly and concisely specify its objectives as a territorial agent, framing these actions within the objectives of sustainable development.

7 REFERENCES

Adobe Inc. (01 de 04 de 2019). Adobe Illustrator. From <https://adobe.com/products/illustrator>

Asamblea General de las Naciones Unidas. (2015). Transformar nuestro mundo: la Agenda 2030 para el Desarrollo Sostenible. From Transformar nuestro mundo: la Agenda 2030 para el Desarrollo Sostenible: https://www.un.org/ga/search/view_doc.asp?symbol=A%2F70%2FL.

Boukraâ, D., i Boussaid, O. (09 de 2009). *Complex Data Warehouse Modelling*. Retrieved March, 19, 2021, from https://www.researchgate.net/profile/Doukifli-Boukraa/publication/228574017_Complex_Data_Warehouse_Modelling/links/549d886e0cf2b803713a787e/Complex-Data-Warehouse-Modelling.pdf

Caiaffa, W. T. (2014). Developing a conceptual framework of urban health observatories toward integrating research and evidence into urban policy for health and health equity. *Journal of urban health : bulletin of the New York Academy of Medicine*, 91(1), 1-16. doi:<https://doi.org/10.1007/s11524-013-9812-0>

Cairo, A. (2016). *The truthful art: Data, charts, and maps for communication*. Estats Units d'Amèrica: New Riders.

Castillo, C. (2015). Developing an academia-based public health observatory: the new global public health observatory with emphasis on urban health at Johns Hopkins Bloomberg School of Public Health. (S. P. Health, Ed.) *Cadernos de saude publica*, 31(supl.1), 286-293. doi:<https://doi.org/10.1590/0102-311X00132914>

Chacon, S., i Straub, B. (2014). *Pro git*. Nova York, Estats Units d'Amèrica: Apress, Berkeley, CA. doi:<https://doi.org/10.1007/978-1-4842-0076-6>

Appendices

Chang, W., Cheng, J., Allaire, J., Xie, Y., i McPherson, J. (20 de 11 de 2020). *shiny: Web Application Framework for R*. From <https://CRAN.R-project.org/package=shiny>

Cheng, J., Sievert, C., Chang, W., Xie, Y., i Allen, J. (2020). *htmltools: Tools for HTML*. From <https://github.com/rstudio/htmltools>

Cofiño, R., Pasarín, M. I., i Segura, A. (2012). ¿Cómo abordar la dimensión colectiva de la salud de las personas? Informe SESPAS 2012. *Gaceta Sanitaria*, 26, 88-93. doi:<https://doi.org/10.1016/j.gaceta.2011.07.030>

Comisión de Estadística de la ONU. (03-06 / 03 / 2020). *Informe sobre el 51er período de sesiones*. From [unstats: https://unstats.un.org/unsd/statcom/51st-session/documents/2020-37-FinalReport-S.pdf](https://unstats.un.org/unsd/statcom/51st-session/documents/2020-37-FinalReport-S.pdf)

Comisión de Estadística de las Naciones Unidas. (7-10 / 03 / 2017). Retrieved March, 19, 2021, from <https://undocs.org/pdf?symbol=es/E/2017/24>

Comisión para Reducir las Desigualdades Sociales en Salud en España. (2012). Propuesta de políticas e intervenciones para reducir las desigualdades sociales en salud en España. *Gaceta Sanitaria*, 26(2), 182-189. doi:[10.1016/j.gaceta.2011.07.024](https://doi.org/10.1016/j.gaceta.2011.07.024)

Creative Commons. (s.f.). CC0. February, 16, 2020, from Creative Commons: <https://creativecommons.org/share-your-work/public-domain/cc0>

Csárdi, G. (2021). *cranlogs: Download Logs from the 'RStudio' 'CRAN' Mirror*. R package version 2.1.1. From <https://r-hub.github.io/cranlogs>

Dahlgren, G., i Whitehead, M. (01 / 1991). Policies and strategies to promote social equity in health. *Institute for Futures Studies, Arbetsrapport*, 14. From https://www.researchgate.net/profile/Goeran-Dahlgren/publication/5095964_Policies_and_strategies_to_promote_social_equity_in_health_Background_document_to_WHO_-

Appendices

[_Strategy_paper_for_Europe/links/569540f808aeab58a9a4d946/Policies-and-strategies-to-promote-so](#)

Dahlgren, G., i Whitehead, M. (2006). *European strategies for tackling social inequities in health: Levelling up Part 2*. WHO Regional office for Europe Copenhagen. From https://www.euro.who.int/__data/assets/pdf_file/0018/103824/E89384.pdf

Daponte Codina, A., Cabrera-León, A., Mateo Rodríguez, I., Espinosa de los Monteros, E., Arroyo-Borrell, E., Saez, M., . . . Bolívar. (2019). *Atlas de los determinantes sociales de la salud en España 2019: evolución y variabilidad entre las Comunidades Autónomas*. Publicaciones EASP.

Demski, H., Garde, S., i Hildebrand, C. (2016). Open data models for smart health interconnected applications: the example of openEHR. *BMC Medical Informatics and Decision Making*, 16(1), 137. doi:<https://doi.org/10.1186/s12911-016-0376-2>

Departament de Salut de la Generalitat de Catalunya. (17 / 03 / 2021a). *Registre de casos de COVID-19 realitzats a Catalunya. Segregació per sexe i municipi*. From Dades Obertes Catalunya: <https://analisi.transparenciacatalunya.cat/Salut/Registre-de-casos-de-COVID-19-realitzats-a-Catalun/jj6z-iyrp>

Departament de Salut de la Generalitat de Catalunya. (16 / 03 / 2021b). *Registre de casos de COVID-19 realitzats a Catalunya. Segregació per sexe i edat*. From Dades Obertes Catalunya: <https://analisi.transparenciacatalunya.cat/Salut/Registre-de-casos-de-COVID-19-realitzats-a-Catalun/qwj8-xpvk>

Departament de Salut de la Generalitat de Catalunya. (16 / 03 / 2021c). *Registre de defuncions per COVID-19 a Catalunya. Segregació per sexe i comarca*. From Dades Obertes Catalunya: <https://analisi.transparenciacatalunya.cat/Salut/Registre-de-defuncions-per-COVID-19-a-Catalunya-Se/uqk7-bf9s>

Appendices

Departament de Salut de la Generalitat de Catalunya. (16 / 03 / 2021d). *Vacunació per al COVID-19 per comarca*. Consultat el 16 / 03 / 2021, a Dades Obertes Catalunya: <https://analisi.transparenciacatalunya.cat/Salut/Vacunaci-per-al-COVID-19-per-comarca/cuwj-bh3b>

Devlin, H., Schenk, T. J., Leynes, G., Lucius, N., Malc, J., Silverberg, M., i Schmeideskamp, P. (2019). *RSocrata: Download or Upload 'Socrata' Data Sets*. From <https://CRAN.R-project.org/package=RSocrata>

ECMA International. (12 / 2017). *ECMA-404 The JSON Data Interchange Syntax*. Consultat el 19 / 03 / 2021, a ECMA International: https://www.ecma-international.org/wp-content/uploads/ECMA-404_2nd_edition_december_2017.pdf

Erb, B., Habiger, G., i Hauck, F. J. (2016). On the Potential of Event Sourcing for Retroactive Actor-Based Programming. *First Workshop on Programming Models and Languages for Distributed Computing(9781450347754)*, 1-5. doi:10.1145/2957319.2957378

European Comission. (sense data). *European Comission*. Consultat el 12 / 01 / 2020, a The European Core Health Indicators (ECHI) shortlist of 88 health indicators identified by policy area: https://ec.europa.eu/health/sites/health/files/indicators/docs/echi_shortlist_by_policy_area_en.pdf

EUROSTAT. (2019). *Key figures on Europe — Statistics illustrated — 2019 edition*. (U. B.—D. Helene Strandell and Pascal WolffEurostat, Ed.) Luxembourg, Luxembourg, Luxembourg: EUROSTAT. doi:10.2785/415908

EUROSTAT. (2020). *Key figures on Europe — Statistics illustrated — 2020 edition*. Luxembourg, Luxembourg, Luxembourg: EUROSTAT. doi:10.2785/894405

Appendices

Extensible Markup Language (XML) 1.0 (Fifth Edition). (26 de 11 de 2008).

Obtingut el 17 de 02 de 2020, de W3C: <https://www.w3.org/TR/xml/>

Federación española de municipios y provincias, Red de entidades locales por la transparencia y participación ciudadan, wolter kluwer. (24 de 04 de 2018). *datos.gob.es*. From Datos abiertos: guía estratégica para su puesta en marcha y conjuntos de datos mínimos a publicar: https://datos.gob.es/sites/default/files/doc/file/guia_datos_abiertos.pdf

Few, S. (2013). *Information Dashboard Design: Displaying data for at-a-glance monitoring*. (B. Pierce, Ed.) El Dorado Hills, CA 95762: Analytics Press.

Fowler, Martin, and Jim Highsmith. "The agile manifesto." *Software Development* 9.8 (2001): 28-35.

Fowler, M. (2004). *UML distilled: a brief guide to the standard object modeling language*. Estats Units d'Amèrica: Addison-Wesley Professional.

Fowler, M. (25 / 01 / 2006). *Focusing on Events*. Consultat el 24 / 03 / 2021, a martinfowler.com:
<https://www.martinfowler.com/eaDev/EventNarrative.html>

Fowler, M. (07 / 02 / 2017). *What do you mean by "Event-Driven"?* Consultat el 24 / 03 / 2021, a martinfowler.com:
<https://www.martinfowler.com/articles/201701-event-driven.html>

Franquet Bonet, A., Plaja Roman, P., Iglesias Hernandez, E., Barceló Rado, M. A., i Saez Zafra, M. (2020). *Indicadors Clau de l'Alt Empordà*. Figueres: Departament de Comunicació de la Fundació Salut Empordà. doi:doi:10.48334/I9109-5826-6199-

Franquet, A., Barceló, M. A., Saez, M., i Plaja, P. (26 de 05 de 2020). *eventr: Create Event Based Data Architectures*. From <https://cran.r-project.org/web/packages/eventr/index.html>

Appendices

Free Software Foundation. (31 / 11 / 2020). <https://www.gnu.org/licenses/gpl-3.0.en.html>. From GNU Operating System: <https://www.gnu.org/licenses/gpl-3.0.en.html>

Fundació Salut Empordà. (2019). *Infografies: Demografia*. Consultat el 01 / 03 / 2021, a Indika Pol de Salut i Social de l'Alt Empordà: <https://www.indikasalut.cat/wp-content/uploads/2021/01/01-poblacio-alt-emporda.pdf>

Generalitat de Catalunya. (2018). *Guia de visualització de dades*. Barcelona: Generalitat de Catalunya. From <http://atenciociudadana.gencat.cat/web/.content/manuals/guia-visualitzacio-dades.pdf>

Generalitat de Catalunya. (20 de 11 de 2020). *Dades obertes: Govern obert*. From Govern obert: <https://analisi.transparenciacatalunya.cat/>

GitHub. (30 de 11 de 2020). From GitHub: <https://github.com/>

Griggs, D., Stafford-Smith, M., Gaffney, O., Rockström, J., Öhman, M. C., Shyamsundar, P., . . . Noble, I. (2013). Sustainable development goals for people and planet. *Nature*, 495, 305-307. doi:<https://doi.org/10.1038/495305a>

Hametner, M., Kostetckaia, M., Setz, I., Bley, S.-J., Gebhard, F., Mayer, C., . . . Steurer, A. (2019). *Sustainable development in the European Union: Overview of progress towards the SDGs in an EU context*. Luxembourg: Publications Office of the European Union,. doi:10.2785/58838

Hasenkopf, C., Flasher, J., Veerman, O., i DeWitt, H. L. (2015). OpenAQ: A Platform to Aggregate and Freely Share Global Air Quality Data. *AGU Fall Meeting Abstracts*, A31D--0097.

Hripcsak, G., Duke, J., Shah, N., Reich, C., Huser, V., Schuemie, M., . . . Ryan, P. (2015). Observational Health Data Sciences and Informatics (OHDSI):

Appendices

Opportunities for Observational Researchers. *Studies in health technology and informatics*, 216, 574-578.

Idescat. (2019a). *Padró municipal d'habitants. Per sexe*. Consultat el 10 / 02 / 2021, a Idescat: <https://www.idescat.cat/pub/?id=pmhin=446igeo=com:02>

Idescat. (27 / 12 / 2019b). *Padró municipal d'habitants. Per sexe*. Consultat el 10 / 02 / 2021, a Idescat: <https://www.idescat.cat/pub/?id=pmhin=446igeo=com%3A02iby=munit=201900>

Idescat. (07 / 05 / 2020a). *Estructura per edats, envelliment i dependència*. Consultat el 01 / 03 / 2021, a Idescat: Indicadors demogràfics i de territori: <https://www.idescat.cat/pub/?id=inddtin=915igeo=com:02#Plegable=geo>

Idescat. (28 / 08 / 2020b). *Mortalitat. Taxes per 100.000 habitants. Per tipus de malaltia, sexe i grups d'edat*. From Institut d'Estadística de Catalunya: <https://www.idescat.cat/pub/?id=aecin=826>

Idescat. (05 / 02 / 2021a). *Població a 1 de gener. Per sexe i edat any a any*. From Institut d'Estadística de Catalunya: <https://www.idescat.cat/pub/?id=pmhin=1180igeo=com:02>

Idescat. (15 / 01 / 2021b). *Taxa bruta de natalitat. Comarques i Aran, àmbits i províncies*. Consultat el 05 / 03 / 2021, a Idescat: <https://www.idescat.cat/pub/?id=aecin=264>

Idescat. (15 / 01 / 2021c). *Naixements. Per edats de la mare. Comarques i Aran, àmbits i províncies*. Consultat el 05 / 03 / 2021, a Idescat: <https://www.idescat.cat/pub/?id=aecin=260>

Indika Pol de Salut i Social de l'Alt Empordà. (01 / 01 / 2020). *Indika Pol de Salut i Social de l'Alt Empordà*. From Indika Pol de Salut i Social de l'Alt Empordà: <http://www.indikasalut.cat/>

Indika Pol de Salut i Social de l'Alt Empordà. (22 de 03 de 2021). *La Natalitat a l'Alt empordà*. Obtingut el 22 de 03 de 2021, de Indika:

Appendices

https://www.indikasalut.cat/wp-content/uploads/2021/03/INFOGRAFIA_natalitat_marc21.pdf

Inkscape Project. (02 / 11 / 2003). Consultat el 22 / 03 / 2021, a Inkscape: <https://inkscape.org>

Institute for Statistics and Mathematics of WU (Wirtschaftsuniversität Wien). (22 / 03 / 2021). From The comprehensive R Archive Network: <https://cran.r-project.org/>

Jacobson, B., i Castillo-Salgado, C. (27 / 12 / 2014). *Providing health intelligence to meet local needs: a practical guide to serving local and urban communities through public health observatories*. From https://apps.who.int/iris/bitstream/handle/10665/152645/9789241508162_eng.pdf;sequence=1

Kalra, D., Beale, T., i Heard, S. (2005). The openEHR foundation. *Studies in health technology and informatics*, 115, 153--173.

Kinding, D., Bridget C., B., i Remington, P. (15 / 06 / 2010). Mobilizing Action Toward Community Health (MATCH): metrics, incentives, and partnerships for population health. *Preventing chronic disease*, 7(4), A68.

Kneale, D., Khatwa, M., i Thomas, J. (2016). *Identifying and appraising promising sources of UK clinical, health and social care data for use by NICE*. (U. C. London, Ed.) London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education. From https://www.researchgate.net/profile/Dylan_Kneale/publication/308661443_Identifying_and_appraising_promising_sources_of_UK_clinical_health_and_social_care_data_for_use_by_NICE/links/57ea6ad408ae91a0c8d3c721/Identifying-and-appraising-promising-sources-of-

Llano, J., García, J., Diego, J., i Recalde, J. (2014). *Aportación de los "Real World Data (RWD)" a la mejora de la práctica clínica*. Madrid, Espanya: Fundación Gaspar Casal. From <https://www.researchgate.net/profile/Juan-Llano->

Appendices

2/publication/303881739_Aportacion_de_los_Real_World_Data_RWD_a_la_mejora_de_la_practica_clinica/links/575a822a08aed884620d43a3/Aportacion-de-los-Real-World-Data-RWD-a-la-mejora-de-la-practica-clinica.pdf

ONU. (2020). Marco de indicadores mundiales para los Objetivos de Desarrollo Sostenible y metas de la Agenda 2030 para el Desarrollo Sostenible.

Open Knowledge Foundation. (17 de 02 de 2021a). *Open Data Commons*. Obtingut el 17 de 02 de 2021, de Open Knowledge Foundation: <https://opendatacommons.org/licenses/pddl/1-0/>

Open Knowledge Foundation. (17 de 02 de 2021b). *Open Data Handbook: What Is Open Data?* Obtingut el 17 de 02 de 2021, de Open Data Handbook: <http://opendatahandbook.org/guide/es/what-is-open-data/>

R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. Vienna, Austria. From <https://www.R-project.org/>

Reeder, B., i David, A. (2016). Health at hand: A systematic review of smart watch uses for health and wellness. *Journal of biomedical informatics*, 63, 269-276. doi:<https://doi.org/10.1016/j.jbi.2016.09.001>

RStudio. (23 / 03 / 2021). *shinyapps.io*. Consultat el 23 / 03 / 2021, a shinyapps.io: <https://www.shinyapps.io/>

Scarpato, N., Alessandra, P., Di Nunzio, L., i Fallucchi, F. (12 / 2017). E-health-IoT universe: A review. *International Journal on Advanced Science, Engineering and Information Technology*, 7. doi:10.18517/ijaseit.7.6.4467

Simsion C, G., i Graham C, W. (2004). *Data modeling essentials* (Third Edition ed.). San Francisco: Elsevier.

Staa, T., Goldacre, B., Gulliford, M., Cassell, J., Pirmohamed, M., Taweel, A., Delaney, B. i Smeeth, L. (2012). Pragmatic randomised trials using routine electronic health records: Putting them to the test. *BMJ (Clinical research ed.)*. 344. e55. [10.1136/bmj.e55](https://doi.org/10.1136/bmj.e55).

Appendices

Stachura, F., i Igras, K. (20 de 11 de 2020). *semantic.dashboard: Dashboard with Fomantic UI Support for Shiny*. From semantic.dashboard: Dashboard with Fomantic UI Support for Shiny: <https://CRAN.R-project.org/package=semantic.dashboard>

Stenberg, D. (20 de 11 de 2020). *CURL*. From CURL: <https://curl.se/>

The Association of British Pharmaceutical Industry. (2011a). *The vision for real world data. Harnessing the opportunities in the UK. Demonstrating value with real world data*. Londres, Regne Unit: ABPI.

The Association of British Pharmaceutical Industry. (2011b). *Demonstrating value with real world data: A practical guide*. Londres, Regne Unit: ABPI.

Whitehead M, D. G. (2006). *Levelling up (part 1): a discussion paper on concepts and principles for tackling social inequities in health*. (W. R. Europe, Ed.) Obtingut el 25 de 01 de 2021, de https://www.who.int:https://www.who.int/social_determinants/resources/leveling_up_part1.pdf

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. (S.-V. N. York, Ed.) From <https://ggplot2.tidyverse.org>

Wickham, H. (2019). *Advanced R*. Boca Raton: CRC Press. doi:<https://doi.org/10.1201/9781351201315>

Wickham, H., Averick, M., Bryan, J., Chang, W., D'Agostino McGowan, L., François, R., Grolemond, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, TL., Miller, E., Bache, SM., Müller, K., Ooms, J., Robinson, D., Seidel, DP., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K. , Yutani, H. (2019). Welcome to the {tidyverse}. *Journal of Open Source Software*, 4(43), 1686. doi:10.21105/joss.01686

Wilkinson, Richard G, Marmot, Michael, World Health Organization. Regional Office for Europe, WHO Centre for Urban Health (Europe) i International Centre for Health and Society. (2003). *Social determinants of health: the*

Appendices

solid facts. (R. W. Marmot, Ed.) Copenhagen: WHO Regional Office for Europe. From <https://apps.who.int/iris/handle/10665/108082>

World Health Organization. (2008). *Closing the gap in a generation: health equity through action on the social determinants of health - Final report of the commission on social determinants of health*. World Health Organization. From http://whqlibdoc.who.int/publications/2009/9789243563701_spa.pdf?ua=1

World Health Organization. (2010). *A conceptual framework for action on the social determinants of health*. World Health Organization. From <https://apps.who.int/iris/handle/10665/44489>

World Health Organization. (2016). *Commission on Social Determinants of Health*,. Obtingut el 27 de 01 de 2021, de World Health Organization: http://www.who.int/social_determinants/thecommission/en/

World Health Organization. (10 / 11 / 2020). *Social determinants of health*. From Social determinants of health: https://www.who.int/health-topics/social-determinants-of-health#tab=tab_1

World Health Organization. (s.f.). *Enfermedades crónicas*. Obtingut el 09 de 12 de 2020, de World Health Organization: https://www.who.int/topics/chronic_diseases/es/

World Health Organization. (s.f.). *Enfermedades no transmisibles*. Obtingut el 09 de 12 de 2020, de World Health Organization: https://www.who.int/topics/noncommunicable_diseases/es/

World Health Organization. (s.f.). *Enfermedades no transmisibles y salud mental*. Obtingut el 09 de 12 de 2020, de World Health Organization: <https://www.who.int/nmh/topics/es/>

APPENDICES

CODE OF THE "POPULATION IN ALT EMPORDÀ" INFOGRAPHIC

```

#
# Codi de la infografia "Població a l'Alt Empordà"
#
# Carregar llibreries
library(tidyverse)
# Descarregar dades "Estructura per edats, envelliment i
dependència"
pag <- xml2::read_html(x =
"https://www.idescat.cat/pub/?geo=com%3A02iid=inddtin=915#Plegable=g
eo")
tab <- rvest::html_table(pag)
i <- tab[[1]][, 1] == 2019
x <- tab[[1]][i, ]

# Obtenir els indicadors d'interés.
PoblacioPediatria <- as.numeric(sub(",", ".", x[1,2], fixed =
TRUE))
PoblacioPediatria <- data.frame(
  category = c('P', 'NP'),
  count     = c(PoblacioPediatria, 100 - PoblacioPediatria)
)

# Gràfic de donut (codi modificat de: https://www.r-graph-
gallery.com/128-ring-or-donut-plot.html)
ggplot(PoblacioPediatria, aes(x = 1, y = Count, fill = c('1',
'2')))) +
  geom_bar(stat = 'identity') +
  coord_polar('y')

# Compute percentages
PoblacioPediatria$fraction = PoblacioPediatria$count /
sum(PoblacioPediatria$count)

# Compute the cumulative percentages (top of each rectangle)
PoblacioPediatria$ymax = cumsum(PoblacioPediatria$fraction)

# Compute the bottom of each rectangle
PoblacioPediatria$ymin = c(0, head(PoblacioPediatria$ymax, n=-1))

# Make the plot

```

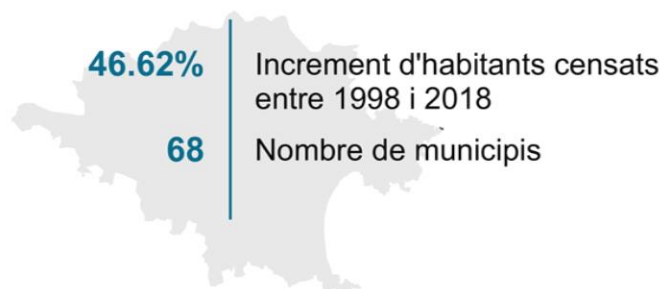
Appendices

```
ggplot(PoblacioPediatria, aes(ymax=ymax, ymin=ymin, xmax=4, xmin=3,
fill=category)) +
  geom_rect() +
  coord_polar(theta="y") + # Try to remove that to understand how
the chart is built initially
  xlim(c(2, 4)) + # Try to remove that to see how to make a pie
chart
  theme_void()
```

POBLACIÓ A L'ALT EMPORDÀ INFOGRAPHIC

POBLACIÓ A L'ALT EMPORDÀ

149.569 habitants censats a l'Alt Empordà



El 16,27% de la població té menys de 15 anys.



El 18,20% de la població té més de 64 anys.



El 2,93% de la població té més de 84 anys.

Altres dades d'interés

El **61,85%** de la població de la comarca de l'Alt Empordà està censada a Figueres (33%), Roses (14%), Castelló d'Empúries (8%) i l'Escala (7%).

La **taxa bruta de natalitat per 1.000 habitants** a l'Alt Empordà l'any 2017 va ser de **9,80** un punt més alt que a **Catalunya** on al mateix període la taxa bruta de natalitat per 1.000 habitants va ser de **8,88** nens/habs.

A **Catalunya** la població amb una **edat inferior als 15** anys és del **15,49%**. La proporció de població amb **més de 64** anys és de **18,70%** i la proporció de persones amb **més de 84** anys és de **3,10%**

Font de les dades: Institut d'Estadística de Catalunya. amb les dades relatives a l'any 2018

Appendices

CODE OF THE “LA NATALITAT A L’ALT EMPORDÀ” INFOGRAPHIC

```
library(tidyverse)

df <- read.csv(
  '../01-data/02-tidy/01-parts-2020.csv',
  sep = ';',
  dec = ',',
  encoding = 'UTF-8'
)

Count <- sum(df$Num)

df <- df[1:10, ]

df$Pais <- factor(df$Pais, levels = rev(df$Pais))

ggplot(df[10:1, ], aes(x = Pais, y = 100 * Num / Count)) +
  geom_bar(aes(y = 100), fill = 'white', stat = 'identity') +
  geom_bar(stat = 'identity') +
  ylim(0, 100) +
  coord_flip()

ggsave(filename = '../04-svg/01-parts.svg')
df <- read.csv(file = '../01-data/01-raw/Consulta Edad Madres Pere
Plaja.csv', sep = ';', encoding = 'UTF-8')

df <- df %>% mutate(data = DNaixement, edat = Edat.mare) %>%
  select(data, edat)

df <- df %>% filter(edat != 227) %>% filter(edat > 3)

ggplot(df, aes(x = edat)) + geom_density() + theme_bw() +
  geom_vline(xintercept = 30) +
  geom_vline(xintercept = c(26, 34))
f <- list.files(path = '../01-data/01-raw/01-partos-por-mes-
idescat/', full.names = TRUE)

x <- lapply(f, read.csv, sep = ';', encoding = 'UTF-8', skip = 5)

names(x) <- 1975:2019

foo = function(y) {
  y <- y[, c(1, 4)]
}
```


Appendices

```
names(y) <- c("Mes", "N")
y %>% filter(Mes != 'Total')

}

z <- purrr::map_df(x, .f = foo, .id = 'Any')
z <- filter(z, Mes != 'No hi consta')

z$Mes <- factor(z$Mes, levels = c("Gener", "Febrer", "Març",
  "Abril", "Maig", "Juny", "Juliol", "Agost", "Setembre", "Octubre",
  "Novembre", "Desembre"))
z$Aux <- 1

ND <- data.frame(
  dia = seq.Date(
    from = as.Date('1975-01-01'),
    to = as.Date('2019-12-31'),
    by = 1
  )
)

ND <- ND %>%
  mutate(year = lubridate::year(dia), month = lubridate::month(dia))
%>%
  group_by(year, month) %>%
  summarise(ND = n(), .groups = 'drop') %>%
  select(ND)

z <- z %>%
  bind_cols(ND) %>%
  group_by(Any) %>%
  # mutate(Mean = mean(N)) %>%
  mutate(NE = sum(N) * ND / sum(ND)) %>%
  mutate(SD = sum(N) * (ND / sum(ND)) * ((sum(ND) - ND)/sum(N))) %>%
  mutate(PR = pnorm(N - NE, 0, 10.84317)) %>%
  mutate(CO = ifelse(PR > .5, '#202d80', '#ffffff'))

ggplot(z, aes(x = Any, y = Mes)) +
  geom_tile(color = '#918cb7', fill = z$CO)

# ggplot(z, aes(x = Any, y = Mes, fill = PR)) +
#   geom_tile(color = '#918cb7', fill = z$CO) +
#   scale_fill_gradientn(colours = c("#faf7ff", "#faf7ff",
# "#918cb7", "#202d80"))

mutate(Color = ifelse(N < NE, '#b5b5b5', '#202d80')) %>%
```

Appendices

```
mutate(Color = factor(Color, levels = c('#202d80', '#b5b5b5'))))

ggplot(z, aes(x = Any, y = 1, group = Mes)) +
  geom_bar(stat = 'identity', colour = '#202d80', fill = z$Color) +
  coord_polar('y')

ggplot(z, aes(x = Mes, y = Any)) +
  geom_tile(color = 'white', fill = z$Color) +
  theme_bw() +
  scale_x_discrete(expand = c(0,0)) +
  scale_y_discrete(expand = c(0,0)) +
  coord_polar()

# w <- z %>% group_by(Mes, Color) %>% summarise(N = n())
#
# ggplot(w, aes(x = Mes, y = N, fill = Color)) + geom_tile(stat =
#'identity') + coord_polar()
f <- list.files(path = '../01-data/01-raw/01-partos-por-mes-
idescat/', full.names = TRUE)

x <- lapply(f, read.csv, sep = ';', encoding = 'UTF-8', skip = 5)

names(x) <- 1975:2019

foo = function(y) {

  y <- y[, c(1, 4)]
  names(y) <- c("Mes", "N")
  y %>% filter(Mes == 'Total')

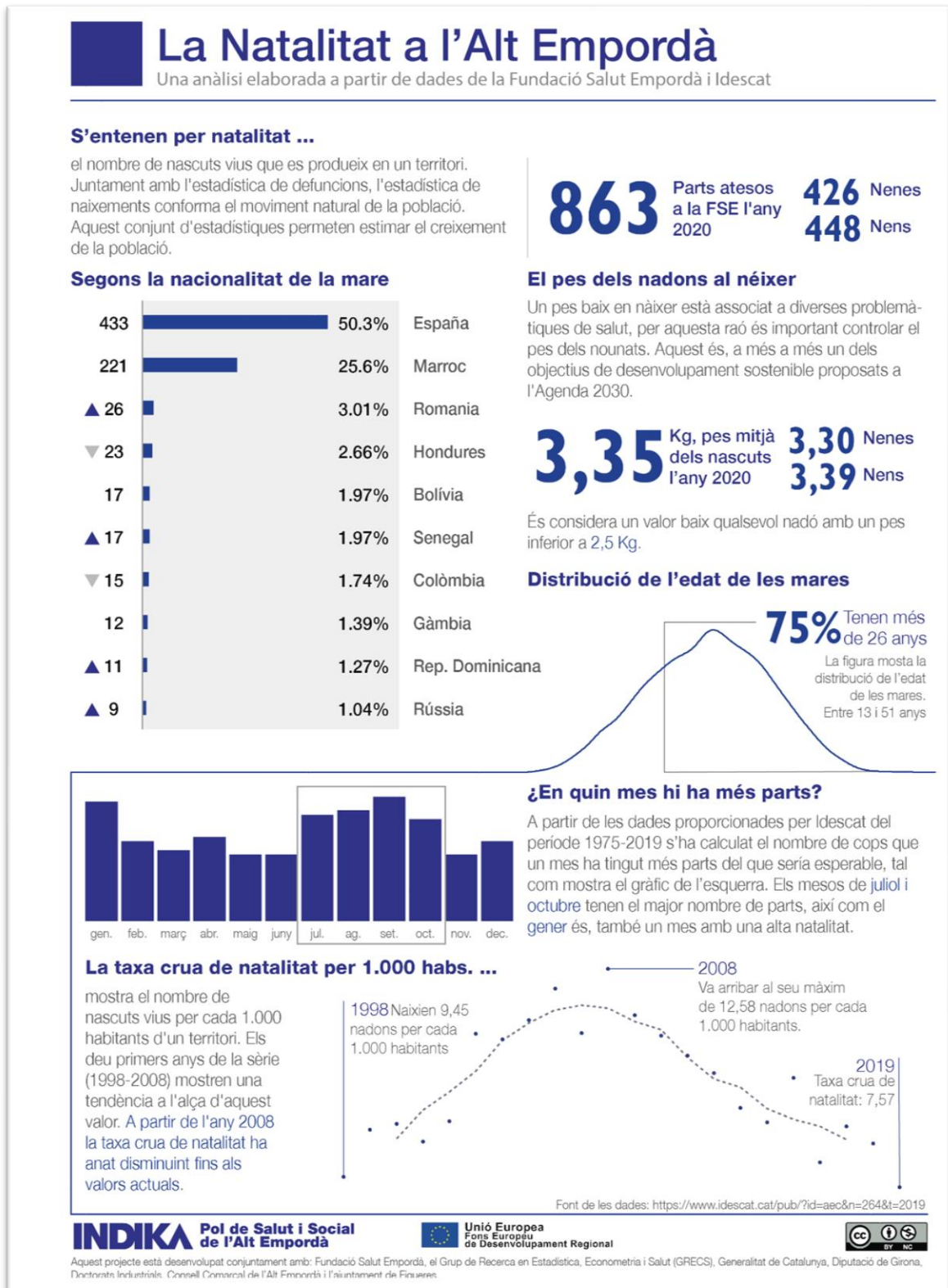
}

z <- purrr::map_df(x, .f = foo, .id = 'Any')
z <- z %>% filter(Any >= 1998)
z$POP <- c(
34546,
35028,
35443,
35777,
36639,
37744,
38824,
40017,
41081,
42586,
44178,
```

Appendices

```
45001,  
45326,  
45189,  
45299,  
44771,  
44578,  
44306,  
44107,  
43995,  
44065,  
44296)  
  
z <- mutate(z, TAXA = 1000 * N / POP)  
z <- mutate(z, MOVINGAVG = forecast::ma(TAXA, order = 5))  
  
z <- data.frame(  
  
  TAXA = c(  
    8.85, 9.74,  
    9.85, 9.51, 9.90, 11.55, 11.45,  
    11.81, 12.41, 11.57, 12.78, 11.90, 11.52, 11.14, 10.81,  
    10.15, 9.88, 10.72, 9.12, 9.80, 9.48, 8.65),  
  Any = c(1998:1999, 2000:2004, 2005:2009, 2010:2014, 2015:2019)  
)  
  
z <- mutate(z, MOVINGAVG = forecast::ma(TAXA, order = 5))  
  
ggplot(z, aes(x = as.integer(Any), y = TAXA)) +  
  geom_line(aes(y = MOVINGAVG), linetype = 'dashed', color =  
'#787586', size = 1) +  
  geom_point(aes(y = TAXA), shape = 21, color = '#202d80', fill =  
'#202d80', size = 3, ) +  
  
  theme_void()
```

LA NATALITAT A L'ALT EMPORDÀ INFOGRAPHIC



Appendices

CODE OF THE "UN ANY DE COVID-19 A L'ALT EMPORDÀ" INFOGRAPHIC

```
# Codi infografia "Un any de COVID-19 a l'Alt Empordà"
## Load packages
library(tidyverse)
library(forecast)
library(RSocrata)

## Download municipality information
URL <- "https://analisi.transparenciaCatalonia.cat/resource/jj6z-
iyrp.json"
### Read credentials
credenciales <- readRDS(file = '01-data/credenciales.rds')

municipalities <-
  read.socrata(URL, app_token = credenciales$app_token, email =
credenciales$email, password = credenciales$password) %>%
  transmute(
    count = as.integer(numcasos),
    sex = as.factor(sexedescripcio),
    result = as.factor(resultatcoviddescripcio),
    data = as.Date(as.character(data)),
    region = comarcacodi
  ) %>% filter(result == "Positiu PCR")

## Catalonia
### Tidy data for Catalonia
catalonia <-
  municipalities %>%
  group_by(data) %>%
  summarise(count = sum(count)) %>%
  filter(data >= '2020-03-09') %>%
  filter(data <= '2021-03-09') %>%
  mutate(freq = 100000 * count / 7727029)

### Plot Moving Average
ggplot(catalonia, aes(x = data, y = ma(freq, 7) %>% as.numeric())) +
  geom_line() +
  theme_void()

## Alt Emporda
### Tidy data
alt <- municipalities %>%
  filter(region == "02") %>%
  group_by(data) %>%
  summarise(count = sum(count)) %>%
```

Appendices

```
filter(data >= '2020-03-09') %>%
filter(data <= '2021-03-09') %>%
mutate(freq = 100000 * count / 142624)

### Plot Moving Average with important dates has vertical lines.
ggplot(alt, aes(x = data, y = ma(freq, 7) %>% as.numeric())) +
  geom_line(colour = '#202d80') +
  geom_point(aes(y = freq), size = .75) +
  theme_void() +
  geom_vline(xintercept = as.Date('2020-03-14')) + # estat d'alarma.
  geom_vline(xintercept = as.Date('2021-02-14')) + # eleccions de
Catalonia.
  geom_vline(xintercept = as.Date('2020-06-21')) +
  geom_vline(xintercept = as.Date('2020-10-25')) +
  geom_vline(xintercept = as.Date('2020-12-24'), color = 'red') +
  geom_vline(xintercept = as.Date('2021-01-07'), color = 'red') +
  geom_vline(xintercept = as.Date('2021-01-14'), color = 'blue')

## Admissions
### Read Data
admissions <- read.csv(file = '01-data/01-raw/02-ingresos-per-
setmana.csv', sep = ';')
admissions_count <- sum(admissions$N)
### Plot admissions
ggplot(admissions, aes(x = 1:nrow(admissions), y = N)) +
  geom_line() +
  theme_void()

## Deaths
### Download Data
URL <- "https://analisi.transparenciaCatalonia.cat/resource/uqk7-
bf9s.json"

deaths <-
  read.socrata(URL,
               app_token = credenciales$app_token,
               email = credenciales$email,
               password = credenciales$password
  ) %>%
  filter(comarcacodi == '02') %>%
  group_by(exitusdata) %>%
  summarise(N = as.integer(numexitus)) %>%
  ungroup() %>%
  filter(exitusdata >= as.Date('2020-03-09')) %>%
  filter(exitusdata <= as.Date('2021-03-09'))
```

Appendices

```
aux <- data.frame(
  exitusdata = seq.Date(
    as.Date('2020-03-01'),
    as.Date('2021-03-01'),
    by = 1
  )
)

deaths <- left_join(aux, deaths) %>%
  replace_na(replace = list(N = 0))

deaths_count <- sum(deaths$N)

monthly_count <- deaths %>%
  mutate(
    year = lubridate::year(exitusdata),
    month = lubridate::month(exitusdata)
  ) %>%
  group_by(year, month) %>%
  summarise(N = sum(N))

ggplot(monthly_count, aes(x = 1:13, y = N)) +
  geom_bar(stat = 'identity', fill = '#202d80') +
  theme_void()

## Infections by age
URL <- "https://analisi.transparenciaCatalonia.cat/resource/qwj8-
xpvk.json"

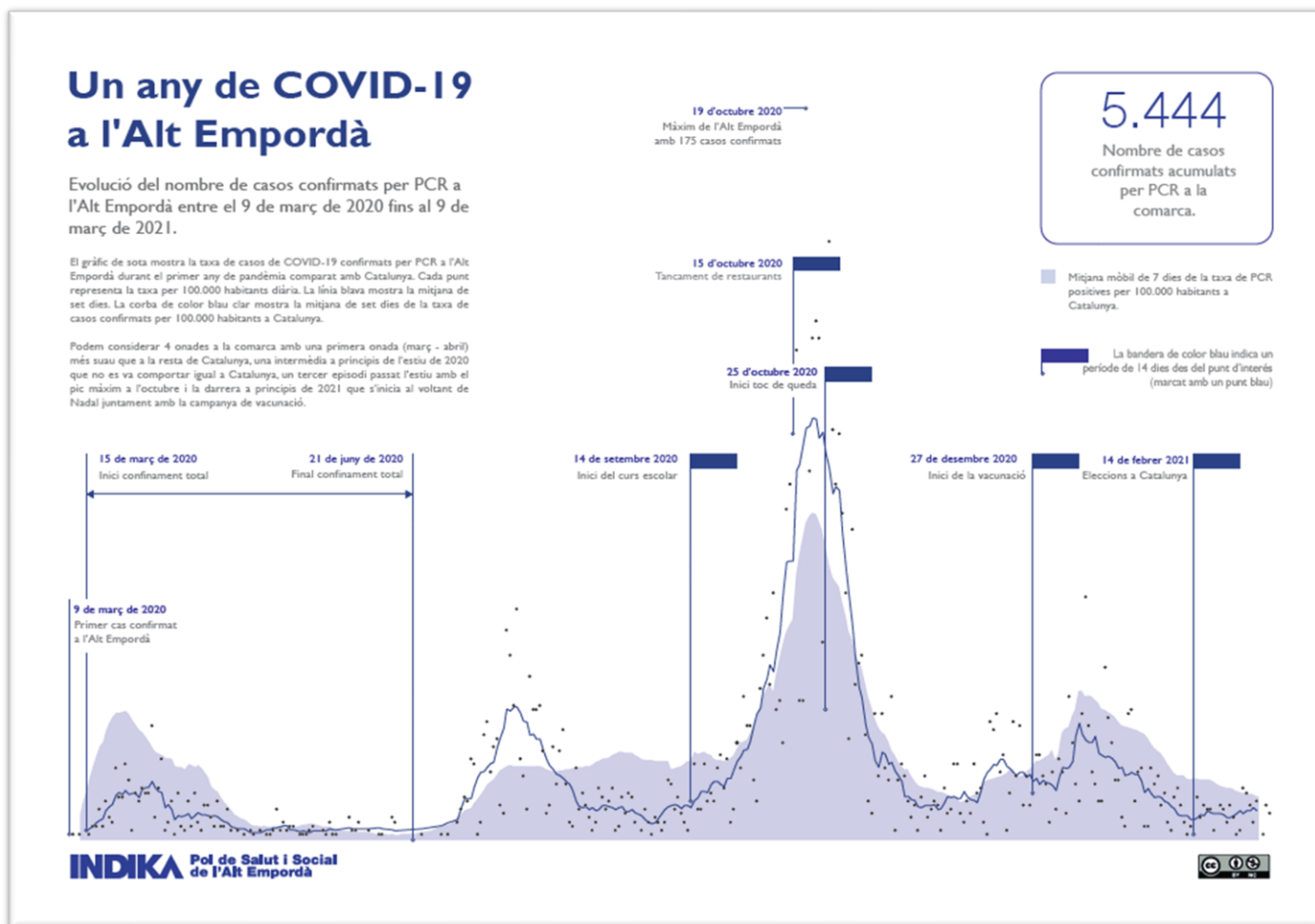
infections <-
  read.socrata(URL,
    app_token = credenciales$app_token,
    email = credenciales$email,
    password = credenciales$password
  ) %>%
  filter(regiosanitariacodi == '6400') %>%
  group_by(edatrang) %>%
  summarise(N = sum(as.integer(numcasos)))

### Populations by age group
pop <- data.frame(
  edatrang = c("0-9", "10-19", "20-29", "30-39", "40-49", "50-59",
    "60-69", "70-79", "80-89", "90+"),
  popu = c(78806, 87126, 84624, 103319, 130224, 110865, 85568,
    59791, 32780, 8685)
```

Appendices

```
)  
  
### Compute 100000 freq  
infections <- left_join(infections, pop) %>%  
  mutate(taxa = 100000 * N / popu)  
  
ggplot(infections[1:10, ], aes(x = edatrang, y = taxa)) +  
  # geom_bar(aes( y = 100000), fill = 'grey', stat = 'identity') +  
  geom_bar(stat = 'identity', fill = '#202d80') +  
  # ylim(0, 100000) +  
  theme_void()
```


UN ANY DE COVID-19 A L'ALT EMPORDÀ INFOGRAPHIC



Nombre d'ingressos hospitalaris totals

La corva d'ingressos ha estat paral·lela a la de casos comarcals (amb petit impasse) ha presentat 4 pics (amb petit desescalatge). Març, juny i octubre de 2020 i Gener de 2021.

Font: Dades de la Fundació Salut Empordà

1.098

Nombre total d'ingressos a la FSE per COVID-19

Nombre de defuncions

El nombre de defuncions s'ha incrementat en el segon semestre analitzat, coincidint amb el 3er i 4t brot.

Font: <https://analisi.transparenciacatalunya.cat/resource/uqk7-bf9s.json>

224

Nombre total de defuncions a l'Alt Empordà.

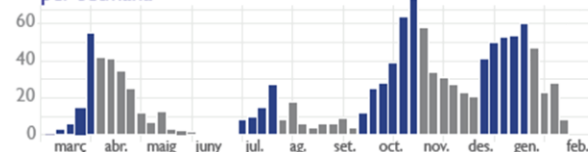
Nombre de casos per edat a Girona

La corva de les edats de la covid esta condicionada per una major detecció en persones grans en el primer període i una disminució de en les edats del segon . probablement afavorit per els screenings associats a àmbit escolar i la millora en les residències.

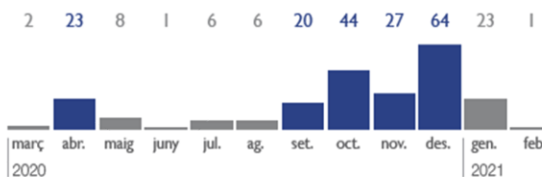
Edat	Núm. casos	%
0-9	3.827	6,24
10-19	7.759	12,60
20-29	7.141	11,60
30-39	8.157	13,30
40-49	10.508	17,10
50-59	8.780	14,30
60-69	5.771	9,41
70-79	3.836	6,25
80-89	3.534	5,76
90+	1.870	3,05

Font: <https://analisi.transparenciacatalunya.cat/resource/qwj8-xpvk.json>

Nombre de nous ingressos hospitalaris per COVID-19 per setmana



Nombre de defuncions mensuals



Estat de la vacunació

A partir del 27 de desembre de 2020 es va iniciar el procés de vacunació a l'Alt Empordà. Al finalitzar el període d'aquesta infografia el nombre de vacunats a l'Alt Empordà és de:

2.411 Nombre total de vacunats amb la segona dosi a l'Alt Empordà

1.69% Percentatge de població vacunada a l'Alt Empordà

Font: <https://analisi.transparenciacatalunya.cat/resource/cuwj-bh3b.json>



CODE OF THE "UN ANY DE COVID-19 A L'ALT EMPORDÀ" INFOGRAPHIC

```
# Codi infografia "Un any de COVID-19 a l'Alt Empordà"
## Load packages
library(tidyverse)
library(forecast)
library(RSocrata)
library(ggmosaic)

## Download municipality information
URL <- "https://analisi.transparenciaCatalonia.cat/resource/jj6z-
iyrp.json"

### Read credentials
credenciales <- readRDS(file = '01-data/credenciales.rds')

municipalities <-
  read.socrata(URL, app_token = credenciales$app_token, email =
credenciales$email, password = credenciales$password) %>%
  transmute(
    count = as.integer(numcasos),
    sex = as.factor(sexedescripcio),
    result = as.factor(resultatcoviddescripcio),
    data = as.Date(as.character(data)),
    region = comarcacodi
  ) %>% filter(result == "Positiu PCR")

## Catalonia
### Tidy data for Catalonia
catalonia <-
  municipalities %>%
  group_by(data) %>%
  summarise(count = sum(count)) %>%
  filter(data >= '2020-03-09') %>%
  filter(data <= '2021-03-09') %>%
  mutate(freq = 100000 * count / 7727029)

### Plot Moving Average
ggplot(catalonia, aes(x = data, y = ma(freq, 7) %>% as.numeric())) +
  geom_line() +
  geom_vline(xintercept = as.Date("2020-03-09")) +
  geom_vline(xintercept = as.Date("2021-03-09")) +
  geom_hline(yintercept = 122.7002) +
  geom_hline(yintercept = 0) +
  theme_void()
```

Appendices

```
## Alt Emporda
### Tidy data
alt <- municipalities %>%
  filter(region == "02") %>%
  group_by(data) %>%
  summarise(count = sum(count)) %>%
  filter(data >= '2020-03-09') %>%
  filter(data <= '2021-03-09') %>%
  mutate(freq = 100000 * count / 142624)

### Plot Moving Average with important dates has vertical lines.
ggplot(alt, aes(x = data, y = ma(freq, 7) %>% as.numeric())) +
  geom_line(colour = '#202d80') +
  geom_point(aes(y = freq), size = 1, shape = 21, fill = 'black') +
  theme_void() +
  geom_vline(xintercept = as.Date('2020-03-14')) + # estat d'alarma.
  geom_vline(xintercept = as.Date('2020-06-21')) +
  geom_vline(xintercept = as.Date('2020-10-15')) +
  geom_vline(xintercept = as.Date('2020-10-25')) + # inici toc de
queda
  geom_vline(xintercept = as.Date('2020-12-24')) +
  geom_vline(xintercept = as.Date('2020-12-27')) + # inici vacunació
  geom_vline(xintercept = as.Date('2021-01-07')) +
  geom_vline(xintercept = as.Date('2021-01-14')) +
  geom_vline(xintercept = as.Date('2021-02-14')) + # eleccions de
Catalonia.
  geom_vline(xintercept = as.Date("2020-03-09"), color = 'red') +
  geom_vline(xintercept = as.Date("2021-03-09"), color = 'red') +
  geom_vline(xintercept = as.Date("2020-09-14")) +
  geom_vline(xintercept = as.Date("2020-09-28")) +
  geom_hline(yintercept = 0)

alt_count <- alt$count %>% sum()

## Admissions
### Read Data
admissions <-
  read.csv(file = '01-data/01-raw/02-ingresos-per-setmana.csv', sep
= ';') %>%
  mutate(week = setmana, year = any) %>%
  select(year, week, NADMIN = N)

admissions_count <- sum(admissions$N)

### Plot admissions
```

Appendices

```
ggplot(admissions, aes(x = 1:nrow(admissions), y = NADMIN)) +
  geom_bar(stat = 'identity') +
  theme_minimal() +
  xlab(element_blank()) +
  ylab(element_blank())

## Deaths
### Download Data
URL <- "https://analisi.transparenciaCatalonia.cat/resource/uqk7-
bf9s.json"

deaths <-
  read.socrata(URL, app_token = credenciales$app_token, email =
credenciales$email, password = credenciales$password) %>%
  filter(comarcacodi == '02') %>%
  group_by(exitusdata) %>%
  summarise(N = as.integer(numexitus)) %>%
  ungroup() %>%
  filter(exitusdata >= as.Date('2020-03-09')) %>%
  filter(exitusdata <= as.Date('2021-03-09'))

aux <- data.frame(
  exitusdata = seq.Date(
    as.Date('2020-03-01'),
    as.Date('2021-03-01'),
    by = 1
  )
)

deaths <- left_join(aux, deaths) %>%
  replace_na(replace = list(N = 0))

deaths_count <- sum(deaths$N)

monthly_count <- deaths %>%
  mutate(
    year = lubridate::year(exitusdata),
    month = lubridate::month(exitusdata)
  ) %>%
  group_by(year, month) %>%
  summarise(N = sum(N))

ggplot(monthly_count, aes(x = 1:12, y = N)) +
  geom_bar(stat = 'identity', fill = '#202d80') +
  theme_void()
```

Appendices

```
## Infections by age
URL <- "https://analisi.transparenciaCatalonia.cat/resource/qwj8-
xpvk.json"

infections <-
  read.socrata(URL,
               app_token = credenciales$app_token,
               email = credenciales$email,
               password = credenciales$password
  ) %>%
  filter(data >= '2020-03-09') %>%
  filter(data <= '2021-03-09') %>%
  filter(regiosanitariacodi == '6400') %>%
  group_by(edatrang) %>%
  summarise(N = sum(as.integer(numcasos))) %>%
  mutate(NF = 100 * N / sum(N)) %>%
  .[-11, ]

ggplot(infections, aes(x = edatrang, y = N)) +
  geom_bar(fill = 'grey', aes(y = rep(100,10)), stat = 'identity') +
  geom_bar(fill = "#202d80", stat = 'identity') +
  coord_flip() +
  theme_void()

# Vaccination
URL <- "https://analisi.transparenciaCatalonia.cat/resource/cuwj-
bh3b.json"

vaccined <- read.socrata(URL, app_token = credenciales$app_token,
                        email = credenciales$email, password = credenciales$password)
vaccined <- vaccined %>%
  filter(comarca == 'ALT EMPORDA') %>%
  filter(data <= as.Date("2021-03-09")) %>%
  filter(!is.na(fabricant)) %>%
  group_by(dosi) %>%
  summarise(N = sum(as.integer(recompte)))
```

KEY FIGURES ON EUROPE

This table shows the review of the indicators available according to their level of territorial breakdown and based on the list of indicators presented in the document *Key Figures On Europe* (EUROSTAT, 2019). Blue indicates that the indicator is available for the level of breakdown stated in the column header, which can be Europe Europa (EU), Spain (ES), Catalonia (CAT), Girona Province (GIR), county of Alt Empordà (ALT), and municipalities of Alt Empordà (MUN).

Capitol	Section	Indicator	Availability					
			EU	ES	CAT	GIR	ALT	MUN
People and society	Population	Population	■	■	■	■	■	■
		Population developments	■	■	■	■	■	■
		Population change	■	■	■	■	■	■
		Fertility rate	■	■	■			

Appendices

		Ageing population						
		Population by citizenship						
		First time asylum application						
	Health	Developments for life expectancy at birth						
		Life expectancy at birth						
		Main causes of death						
		Deaths from diseases of the circulatory system						
	Education	Early leavers from education and training						

Appendices

		Learning two or more foreign languages	■	■	■			
		Young people neither in employment nor in education and training	■	■	■			
		Fields of study for university graduates	■	■	■			
	Labour market	Labour force composition	■					
		Employment rate	■	■	■	■	■	■
		Unemployment rate development	■	■	■	■	■	■
		Unemployment rate	■	■	■	■	■	■

Appendices

		Gender pay gap						
Living condition		People at risk of poverty or social exclusion						
		People at risk of poverty or social exclusion						
		Inability to face unexpected financial expenses						
		Inability to afford to pay for a one-week annual holiday away from home						
Digital society		Internet usage						
		Internet activities						

Appendices

Economy and business	Economy and finance	GDP						
		Real change in GDP						
		GDP per capita						
	Prices	Inflation rate						
		Household budget structure						
		Comparative price levels						
	Government finance	General government deficit/surplus and debt						
		Structure of general government expenditure						

Appendices

		International trade in goods with non-member countries						
		Top 10 partners for EU International trade in goods						
	International trade	International trade in goods with Member States						
		International trade in services with non-member countries						
		Top 10 partners for EU International trade in services						

Appendices

		Change in the structure of value added						
		Gross value added by sector						
	Business	Change in the structure of employment						
		Employment by sector						
		Non-financial business economy						
		Value added in the non-financial business economy						

Appendices

		Developments for industrial output						
	Research and development	Gross domestic expenditure on R i D						
	Tourism	Nights spent in EU tourist accommodation						
		Top 20 EU regions in terms of nights spent in tourist accommodation						
		Nights spent in tourist accommodation						
Environment and natural resources	Transport	Passenger cars						
		Age of passenger cars						

Appendices

		Air passenger transport					
		Top 20 EU airport					
		Inland freight transport by type of transport					
		Top 20 EU short sea shipping ports					
	Energy	Origin of energy imports					
		Energy dependency rate					
		Energy from renewable energy source					
		Natural gas and electricity prices					

Appendices

		Structure of final energy consumption						
		Final energy consumption by end use						
	Environment	Employment in the environmental economy						
		Environmental tax revenue						
		Greenhouse gas emissions						
		Greenhouse gas emissions by source sector						

Appendices

		Municipal waste generation						
		Municipal waste treatment methods						
	Agriculture	Top 5 EU Member States with the largest production for selected agricultural products						
		Organic crop area						
	Fisheries	EU Member States with the largest fish catches						
	Forestry	Area of forest and other wooded land						

Appendices

LIST OF KEY INDICATORS FOR ALT EMPORDÀ

		Availability			
		CAT	GIR	ALT	URL
People and population	Population				http://www.idescat.cat/pub/?id=pmhin=446igeo=com:02iby=mun
	Evolution of the population				http://www.idescat.cat/pub/?id=ep
	Change in the population				http://www.idescat.cat/pub/?id=ep
	Gros birth rate				http://www.idescat.cat/pub/?id=aecin=264

Appendices

		Population aging				http://www.idescat.cat/pub/?id=proj
		Population: by gender and five-year period				http://www.idescat.cat/pub/?id=pmhin=9548igeo=com:02ilang=es
		Population by citizenship				http://www.idescat.cat/pub/?id=pmhin=679iby=munigeo=com%3A02#Plegable=geo
	Health	Evolution of life expectancy at birth				http://www.idescat.cat/pub/?id=ievin=8619igeo=prov:17#Plegable=geo
		Main cause of death				Dades de la Fundació Salut Empordà
	Education	Rate of schooling at age 17				http://www.idescat.cat/pub/?geo=com%3A02iid=intpobrin=8233#Plegable=geo
	Workforce	Composition of the workforce				http://www.idescat.cat/pub/?id=censphin=270iby=com

Appendices

		Evolution of the registered unemployed				http://observatorideltreball.gencat.cat/ObservatoriDeITreball/servlet/mstrWeb
		Registered rate of unemployment				https://observatoritreball.gencat.cat/ca/inici
	Life Conditions	Income inequalities according to tax sources				http://www.idescat.cat/pub/?id=intpobrin=8228iby=com
		Percentage of incomes lower than 60% of the average				http://www.idescat.cat/pub/?id=intpobrin=8228iby=com
		Gap in incomes below 60% of the average				http://www.idescat.cat/pub/?id=intpobrin=8228iby=com
		Gender pay gap				http://www.idescat.cat/indicadors/?id=anualsin=10405

Appendices

	Digital Society	Us of Internet				http://www.idescat.cat/pub/?id=aecin=621
		Activity on internet				http://www.idescat.cat/pub/?id=ticll19in=3.3.1
Economy and Business	Economy and Finances	Gross interior product				http://www.idescat.cat/pub/?id=pibcin=8276iby=comilang=es
		Rate of variation of the real GNP				http://www.idescat.cat/pub/?geo=com%3A02iid=pibcin=8276#Plegable=geo
		GNP per capital Alt Empordà				http://www.idescat.cat/pub/?id=pmhin=1181igeo=com:02ilang=es
	Prices	Household spending				http://www.idescat.cat/pub/?id=edclin=9493
	Finance of the govern	Deficit/surplus and government debt				www.seu-e.cat/ca/web/ccaltemporda/govern-obert-i-transparencia/gestio-economica/gestio-economica/endeutament

Appendices

		Structure of public spending				www.seu-e.cat/ca/web/ccaltemporda/govern-obert-i-transparencia/gestio-economica/pressupost/liquidacio-del-pressupost/despeses-per-programa
	Businesses	Changes in the structure of gross added value				http://www.idescat.cat/pub/?geo=com:02iid=pibcin=8277#Plegable=geo
		Employed population by branch of activity				http://www.idescat.cat/pub/?id=aecin=312
	Tourism	Travellers in hotel establishments				Ask IDESCAT
		Overnight stays in hotel establishments				Ask IDESCAT
		Degrees of occupation in hotel establishments				Ask IDESCAT

Appendices

Environment and resources natural	Transport	Vehicle park by type			http://www.idescat.cat/pub/?id=aecin=588
		Age of vehicle park			Ask DGT
	Energy	Final structure of electrical energy consumption			Ask Institut Català d'Energia
		Final structure of natural gas consumption			Ask Institut Català d'Energia
	Environment	Municipal waste generation			http://www.idescat.cat/pub/?id=resmcin=6997igeo=com:02
		Type of waste			http://www.idescat.cat/pub/?id=resmcin=7000igeo=com:02 http://www.idescat.cat/pub/?id=resmcin=6998igeo=com:02

Appendices

	Agriculture	Land use by crop according to type			http://www.idescat.cat/pub/?id=aecin=202
	Forestry	Land use			http://www.idescat.cat/pub/?id=aecin=202



DOCTORAL THESIS

**Mètodes estadístics per gestionar i
analitzar *Real World Data*.**

**Desenvolupament d'un observatori de
salut a l'Alt Empordà**

Aquesta tesi conté un annex amb els codis utilitzats i les infografies obtingudes com a resultat. La tesi es presenta en anglès i català amb la intenció de facilitar-ne la difusió.

Álvaro Franquet Bonet

2021

Programa de Doctorat en Biologia Molecular, Biomedicina i Salut

Sota la direcció de:

Dr. Maria Antònia Barceló Rado i Dr. Pere Plaja Roman

Tesi lliurada per obtenir el títol de doctor per la Universitat de Girona



La **Prof. Dr. Maria Antònia Barceló Rado**, de la Universitat de Girona i membre del Grup de Recerca en Estadística, Econometria i Salut (GRECS) i el **Dr. Pere Plaja Roman**, de la Universitat de Girona i director d'Innovació i Recerca de la Fundació Salut Empordà

CERTIFIQUEN:

Que el treball titulat "Statistical methods to manage and analyse Real World Data. Development of a health observatory in Alt Empordà", que presenta Alvaro Franquet Bonet per a l'obtenció del títol de doctor, ha estat realitzat sota la seva direcció i que compleix els requisits per poder optar a Menció Internacional i a Menció Industrial.

I, perquè així consti i tingui els efectes oportuns, signem aquest document.

Girona, 2 de juliol de 2021

AGRAÏMENTS

Gràcies als directors de la tesi. A **Pere** per la teva escolta i confiança en les propostes que han anat sorgint al llarg del procés. **Maria Antònia** gràcies per la teva paciència i dedicació.

Gràcies a la gent de la **Fundació Salut Empordà**, especialment a **Eneida** i a **Lluís** per tots els moments al despatx, als nostres dinars de menú on hem rigut tant i a tota la feina que han fet per tal que el projecte tingui finançament. Moltes gràcies a **Berta** per les correccions i per assegurar-se que les coses surten bé. Gràcies a **Jordi, Sergi, Ferran** i **Forment** pels riures al despatx.

Gràcies a la gent del **GRECS** que, tot hi haver estat poc pèl despatx, m'heu fet sentir molt a gust. Sobretot a **Marc** per la seva dedicació i a **Marta** per respondre, sempre amb un somriure, a tots els meus dubtes.

Thanks **Ruut** and **Ivonne** for the wonderful **Erasmus University Rotterdam** experience, where I had the pleasure to participate in the elaboration of the paper "HOW TO CREATE THE MOST COMPLETE LISTING OF OVERALL HAPPINESS IN NATIONS: Methods for estimating average life-satisfaction in nations/years where questions on that matter have not been asked" with you.

Gràcies a la gent de la **Universitat Politècnica de Valencia** per acceptar-me en els cursos d'estadística Bayesiana i per l'experiència d'estar amb ells durant el temps que va durar el curs.

*Gracias a la gente de la Escuela Andaluza de Salud per cuidarnos tan bien durante las reuniones y congreso. A **Andrés** por compartir su experiencia en la construcción de observatorios de salud.*

Finalment, gràcies a totes les persones meravelloses – **amics i família** – que m'heu ajudat de moltes i diverses formes a seguir per aquest camí i aconseguir el que m'havia proposat. Gràcies **Albert** i **Katerina** per acollir-me i fer-me sentir com a casa. Gràcies **Carmesina**, per *estar* en tot moment i de totes les maneres possibles, aquesta tesi no seria possible sense la teva paciència. I gràcies

mama, Marc i Pili per acompanyar-me tant i tant bé aquests 30 anys. Finalment, als que ja no hi són però que de ben segur s'alegrarien tant com els que sí que hi són.

Moltíssimes gràcies a tots

!

Álvaro

FINANÇAMENT

Aquesta tesi doctoral va comptar amb el suport de la Convocatòria 'Doctorats Industrial 2018 (DI 2018)'. Primera Resolució parcial definitiva de la convocatòria de doctorats industrials (DI) 2018 del president de la Comissions Executives d'Ajuts Universitaris i de Recerca, 10 de setembre de 2018. Directors de tesis: Maria Antònia Barceló (GRECS, UdG) i Pere Plaja (Fundació Salut Empordà).

També va comptar amb el suport del Fons Europeu de Desenvolupament Regional (FEDER) associat a la convocatòria pública de subvencions del Servei de Programes Europeus, publicada al Butlletí Oficial de la Província de Girona núm. 94 de 18 de maig de 2016.

Està associat a la part de desenvolupament del Pol de Salut i Social de l'Alt Empordà, promogut per la Fundació Salut Empordà i dins del marc del projecte «Girona, regió saludable» (Projecte Europeu de Col·laboració Territorial, PECT), en el qual també participen la Diputació de Girona, el Dipsalut i l'IdibGi

LLISTAT D'ABREVIATURES

API *Application Programming Interface*

AQuAS *Agència de Qualitat i Avaluació Sanitàries de Catalunya*

CDM *Common Data Model*

COVID-19 *Coronavirus Disease 2019*

CRAN *Comprehensive R Archive Network*

CSV *Comma-separated value*

cURL *Client Uniform Resource Locator*

DGT *Dirección general de tráfico*

Dipsalut *Organisme de Salut Pública de la Diputació de Girona*

DM2 *Diabetes Mellitus tipo II*

ECAP *Estació clínica d'atenció primària*

EDA *Event Driven Architecture*

ELISA *Enzyme-linked immunoassay*

EUROSTAT *Oficina Europea de Estadística*

FSE *Fundació Salut Empordà*

GPL *General Public License*

GRECS *Grup de Recerca en Estadística, Econometria i Salut*

HTML *HyperText Markup Language*

Llistat d'abreviatures

IAEG-SDG *Inter-agency and Expert Group on Sustainable Development Goals Indicators*

Idescat *Institut d'Estadística de Catalunya*

IdiBGi *Institut d'Investigació Biomèdica de Girona Dr. Josep Trueta*

INE *Instituto Nacional de Estadística*

JSON *JavaScript Object Notation*

OBSA *Observatorio de Salud de Asturias*

ODS *Objetivos de desarrollo sostenible*

OHDSI *Observational Health Data Sciences and Informatics*

OKF *Open Knowledge Foundation*

OMS *Organicació Mundial de la Salut*

ONU *Organització de les Nacions Unides*

OpenEHR *Open Electronic Health Records*

PADRIS *Programa d'analítica de dades per a la recerca i la innovació en salut*

PCR *Polymerase chain reaction*

PECT *Projecte d'Especialització i Competitivitat Territorial*

RCA *Registre Central d'Assegurats*

RCT *Randomized clinical trial*

RWD *Real World Data*

RWE *Real World Evidence*

SDH *Social Determinants of Health*

Llistat d'abreviatures

UCI *Unitat de Cures Intensives*

UdG *Universitat de Girona*

URL *Uniform Resource Locator*

W3C *World Wide Web Consortium*

WHO *World Health Organization*

WU *Wirtschaftsuniversität Wien*

XML *Extensible Markup Language*

ÍNDEX DE FIGURES

Figura 1 Model de determinants de salut de Dahlgren i Whitehead. Font: Dahlgren i Whitehead (1991).....	3
Figura 2 Marc conceptual dels Determinants Socials de la Salut. Font: Comisión para reducir las desigualdades en salud en España (2012).....	4
Figura 3 Model de treball proposat per l'OBSA. Font: Elaboració pròpia.	12
Figura 4 Model del procés de creació d'una infografia. Font: Elaboració pròpia.	21
Figura 5 Esbós d'una infografia. Font: Elaboració pròpia.	23
Figura 6 Esquema d'objecte JSON. Font: Elaboració pròpia a partir de https://www.json.org/json-en.html	30
Figura 7 Esquema de definició d'una matriu mitjançant JSON. Font: Elaboració pròpia a partir de https://www.json.org/json-en.html	31
Figura 8 Esquema d'una etiqueta en format XML. Font: Elaboració pròpia	32
Figura 9 Esquema d'un element en format XML. Font: Elaboració pròpia.	32
Figura 10 Esquema d'un element en format XML amb atributs. Font: Elaboració pròpia.	33
Figura 11: Diagrama de flux d'una arquitectura basada en esdeveniments. Font: Elaboració pròpia	43
Figura 12 Tres esbossos per a la infografia "Població a l'Alt Empordà". Elaboració pròpia.	50
Figura 13 Infografia "Població a l'Alt Empordà" versió 2019. Elaboració pròpia.	51
Figura 14 Esbós de la infografia "Natalitat a l'Alt Empordà" i infografia final. Elaboració pròpia.	52
Figura 15 Esbós de la infografia "Un any de COVID-19 a l'Alt Empordà". Font: Elaboració pròpia.	53

Índex de figures

Figura 16 Prototip de la infografia "Un any de COVID-19 a l'Alt Empordà". Font: Elaboració pròpia.	54
Figura 17 Infografia "Un any de COVID-19 a l'Alt Empordà". Font: Elaboració pròpia.	54
Figura 18 Captura de pantalla del fitxer de dades diàries de Covid-19 per comarca del Departament de Salut. Elaboració pròpia.	55
Figura 19 Aplicació "Situació de la COVID-19 a l'Alt Empordà". Font: Elaboració pròpia.	61
Figura 20 Portada d'Indicadors Clau de l'Alt Empordà 2020. Font: Elaboració pròpia	63
Figura 21 Descàrregues de la llibreria eventr des del 2020-05-26 fins al 2021-06-04. Font: Elaboració pròpia a partir de dades obtingudes mitjançant la llibreria cranlogs.....	88
Figura 22 Pàgina inicial del repositori de dades d'Indika. Font: Elaboració pròpia.	89

ÍNDEX DE TAULES

Taula 1 Llistat d'observatoris analitzats per a la definició d'Indika.	9
Taula 2 Atributs d'atenció prèvia de la percepció visual.	24
Taula 3 Llicències aptes per a dades obertes.	28
Taula 4 Classificació de dades obertes mitjançant el sistema d'estrelles.....	29
Taula 5 Meta informació dels indicadors al repositori.....	41
Taula 6 Exemple d'ocurrències de tres esdeveniments. Font: Elaboració pròpia.	44
Taula 7 Dades diàries de COVID-19 per comarca	56
Taula 8 Estructura de la taula “Registre de casos de COVID-19 realitzats a Catalunya”	58
Taula 9 Trànsit de la pàgina web d'Indika entre els mesos de Gener i Juny....	64
Taula 10 Nombre de descàrregues mensuals del paquet eventr des del repositori del CRAN. Font: Elaboració pròpia a partir de les dades obtingudes mitjançant el paquet cranlogs.....	88

ABSTRACT

The objective of this thesis is to explore in greater depth the process of creating a public health observatory and the processing of the information needed to take territorial decisions in the county of Alt Empordà. The outcome is the creation of Indika, the Health and Social Pole of the Alt Empordà. Indika is a public health agency that acts as the territory's health and well-being agent. The objective of Indika is to generate knowledge about the impact of social determinants of health in Alt Empordà. To do so, it uses a working framework based on three points: to inform, to discuss, and to act. This working framework served to produce the visual information solutions, including infographics, collections of indicators, and a web application whose purpose is to portray the existing problems in the territory to start debate with the different political and social agents to take decisions on improvement initiatives.

Apart from creating a public health observatory, two parallel lines of action have also been developed in this thesis: the creation of public structures to store information based on events as a way of structuring medical information. These two lines of investigation have resulted in the creation of the "Indika Data Repository", an information repository about the county of Alt Empordà, using the collaborative platform GitHub and the implementation of the eventr package in R programming language, and currently published in the Comprehensive R Archive Network. Eventr is a package whose purpose is to facilitate the implementation of architectures based on events.

Key words: public health observatory, data repository, Real World Data, architecture based on events.

RESUM

Aquesta tesi té com a objectiu aprofundir en el procés de creació d'un observatori de salut pública i el tractament de la informació necessària per a la presa de decisions territorials a la comarca de l'Alt Empordà.

El resultat ha estat la creació d'Indika pol de salut i social de l'Alt Empordà. Indika és un observatori de salut pública que actua com a agent de salut i benestar del territori. L'objectiu d'Indika és generar coneixement sobre l'impacte dels determinants socials de salut a l'Alt Empordà. Per a això fa ús d'un marc de treball basat en tres punts: informar, conversar i actuar.

Aquest marc de treball s'ha concretat en la generació de visualitzacions d'informació com infografies, col·leccions d'indicadors i una aplicació web amb l'objectiu de donar a conèixer les problemàtiques del territori i iniciar així un debat amb els diferents actors polítics i socials que permeti concretar accions de millora.

En aquesta tesi, a més de la creació d'un observatori de salut pública, s'han desenvolupat dues línies d'investigació paral·leles que són: la creació d'estructures públiques d'emmagatzematge d'informació i l'anàlisi de l'arquitectura basada en esdeveniments com a forma d'estructurar la informació mèdica.

Aquestes dues línies d'investigació s'han concretat en la creació de "Indika Data Repository", un repositori d'informació sobre la comarca de l'Alt Empordà a través de la plataforma col·laborativa GitHub i la implementació de la llibreria `eventr` en el llenguatge de programació R publicada actualment al Comprehensive R Archive Network. `eventr` és una llibreria que té com a objectiu facilitar la implementació d'arquitectures basades en esdeveniments.

Paraules clau: observatori de salut pública, repositori de dades, Real World Data, arquitectura basada en esdeveniments.

RESUMEN

Esta tesis tiene como objetivo profundizar en el proceso de creación de un observatorio de salud pública y el tratamiento de la información necesaria para la toma de decisiones territoriales en la comarca del Alt Empordà.

El resultado ha sido la creación de Indika pol de salut i social de l'Alt Empordà. Indika es un observatorio de salud pública que actúa como agente de salud y bienestar del territorio. El objetivo de Indika es generar conocimiento sobre el impacto de los determinantes sociales de salud en el Alt Empordà. Para ello hace uso de un marco de trabajo basado en tres puntos: informar, conversar y actuar.

Este marco de trabajo se ha concretado en la generación de visualizaciones de información como infografías, colecciones de indicadores y una aplicación web con el objetivo de dar a conocer las problemáticas del territorio e iniciar así un debate con los diferentes actores políticos y sociales que permita concretar acciones de mejora.

En esta tesis, además de la creación de un observatorio de salud pública, se han desarrollado dos líneas de investigación paralelas que son: la creación de estructuras públicas de almacenamiento de información y el análisis de la arquitectura basada en eventos como forma de estructurar la información médica.

Estas dos líneas de investigación se han concretado en la creación de “Indika Data Repository”, un repositorio de información sobre la comarca del Alt Empordà a través de la plataforma colaborativa GitHub y la implementación de la librería eventr en el lenguaje de programación R publicada actualmente en el Comprehensive R Archive Network. Eventr es una librería que tiene como objetivo facilitar la implementación de arquitecturas basadas en eventos.

Palabras clave: observatorio de salud pública, repositorio de datos, Real World Data, arquitectura basada en evento

INDEX

AGRAÏMENTS.....	III
FINANÇAMENT	VII
LLISTAT D'ABREVIATURES	IX
ÍNDEX DE FIGURES	XIII
ÍNDEX DE TAULES	XV
ABSTRACT	XVII
RESUM	XIX
RESUMEN	XXI
INDEX	XXIII
1 Antecedents	1
1.1 Determinants socials de la salut.....	2
1.2 Objectius de desenvolupament sostenible	6
1.3 Observatori de salut pública.....	8
1.4 Real world data	13
1.5 Models de dades basats en esdeveniments	17
1.6 Visualització d'informació	19
1.6.1 Característiques d'una visualització.....	20
1.6.2 Metodologia en el procés de creació d'una infografia.....	21
1.6.2.1 Estratègia.....	21
1.6.2.2 Dades.....	22
1.6.2.3 Disseny	22
1.6.3 Principis de la visualització d'informació	23
1.7 Open Data.....	26
1.7.1 Sistema de classificació de dades obertes	29

Index

1.7.2	La Notació d'Objectes JavaScript	30
1.7.3	El Llenguatge de marcatge extensible	31
2	Hipòtesis i Objectius	35
2.1	Hipòtesis	35
2.2	Objectius	35
3	Mètodes	37
3.1	Disseny d'un repositori d'indicadors de salut, socioeconòmics i mediambientals.....	37
3.1.1	Identificació de requeriments	37
3.1.1.1	Revisió d'indicadors	39
3.1.1.2	Requisits tècnics	39
3.1.2	Disseny i arquitectura del repositori.....	40
3.2	Construcció d'un repositori d'informació sanitària, socioeconòmica i mediambiental de l'Alt Empordà	40
3.2.1	API.....	42
3.3	Explotació de la base de dades integrada.....	42
3.3.1	Indicadors Clau de l'Alt Empordà.....	42
3.4	Implementació d'una llibreria de R per al modelatge de dades basades en esdeveniments.....	43
3.4.1	Terminologia.....	43
3.4.2	Implementació	45
4	Resultats	49
4.1	Indika pol de salut i social de l'Alt Empordà	49
4.1.1	Infografies.....	49
4.1.1.1	Població a l'Alt Empordà	49
4.1.1.2	infografia "La Natalitat a l'Alt Empordà".....	51
4.1.1.3	Infografia "Un any de COVID-19 a l'Alt Empordà"	52

Index

4.1.2	Aplicació sobre la situació de la COVID-19 a l'Alt Empordà	54
4.1.3	Indicadors Clau de l'Alt Empordà.....	62
4.1.4	Pàgina web del projecte Indika.....	63
4.1.5	Altres mitjans	64
4.2	Llibreria eventr de R.....	64
4.2.1	Estructura de l'individu.....	65
4.2.2	Crear les especificacions.....	65
4.2.3	Crear els manipuladors.....	67
4.2.4	Crear el repartidor.....	69
4.2.5	Crear les ocurrencies dels esdeveniments	70
4.2.6	Executar els esdeveniments.....	71
4.2.7	Avantatges de l'ús d'EDA	72
4.2.7.1	Obtenir el registre d'auditoria	72
4.2.7.2	Testejar els esdeveniments.....	73
4.2.7.3	Esdeveniments retroactius.....	75
4.2.7.4	Estats alternatius.....	76
4.2.8	Aplicació a un cas real.....	77
4.2.8.1	Estructura de l'individu	78
4.2.8.2	Crear les especificacions dels esdeveniments.....	79
4.2.8.3	Crear els manipuladors	80
4.2.8.4	Crear el repartidor	82
4.2.8.5	Crear les ocurrencies dels individus.....	82
4.2.8.6	Executar els esdeveniments per a cada individu	84
4.2.8.7	Obtenir l'Audit Trail.....	86
4.3	Repositori de dades Indika.....	89
5	Discussió.....	91

Index

5.1	Indika pol de salut i social de l'Alt Empordà	91
5.2	Repositori de dades Indika	94
5.3	Eventr	95
6	Conclusions i treball futur	97
7	Referències	99
	Annexos	111
	<i>Codi de la infografia "Població a l'Alt Empordà"</i>	111
	Infografia "Població a l'Alt Empordà"	113
	<i>Codi de la infografia "Natalitat a l'Alt Empordà"</i>	114
	Infografia "Natalitat a l'Alt Empordà"	118
	<i>Codi de la infografia "Un any de COVID-19 a l'Alt Empordà"</i>	119
	Infografia "Un any de COVID-19 a l'Alt Empordà"	123
	<i>Codi de la infografia "Un any de COVID-19 a l'Alt Empordà"</i>	125
	<i>Key Figures on Europe</i>	129
	Llistat dels indicadors clau de l'Alt Empordà	140
	Recull de premsa	147

1 ANTECEDENTS

Els antecedents d'aquesta tesi s'emmarquen per una banda en el projecte Europeu de col·laboració territorial (PECT) "Girona, regió saludable", del qual en formen part, la Fundació Salut Empordà (FSE), la diputació de Girona a través de l'Organisme de Salut Pública de la Diputació de Girona (Dipsalut) i l'Institut d'Investigació Biomèdica de Girona (IdIBGi). Per altra banda, en un projecte de doctorat Industrial en el qual col·laboren el Grup de Recerca en Estadística, Econometria i Salut (GRECS) de la Universitat de Girona (UdG) i la FSE, el qual conclou amb la presentació d'aquesta tesi.

La Fundació Salut Empordà (FSE) és una fundació sense ànim de lucre. La seva missió és prestar serveis d'atenció integral en l'àmbit de la salut i de la dependència, orientada a les necessitats de la població de la comarca gironina de l'Alt Empordà. La FSE proporciona serveis de salut d'atenció primària (àrea bàsica de salut de l'Escala), d'atenció especialitzada (Hospital de Figueres) i d'atenció socio sanitària (Centre Sociosanitari Bernat Jaume).

En els últims anys i com a resultat de l'impuls i la voluntat d'algunes entitats entre les quals hi ha la FSE, s'han desenvolupat diverses iniciatives destinades a disposar d'un major coneixement sobre la realitat de la comarca de l'Alt Empordà en relació amb el seu estat de salut i els determinants socials que la condicionen (Dahlgren i Whitehead, 1991) prenent com a eix vehicular els Objectius de Desenvolupament Sostenible (ODS) (Griggs, et al., 2013).

La comarca de l'Alt Empordà està formada per seixanta-vuit municipis i una població total de 141.339 habitants (Idescat, 2019a), 46.654 (33%) dels quals empadronats a Figueres (Idescat, 2019b), la capital de comarca i la seu física de la fundació, de l'hospital i del centre socio sanitari. En total l'entitat proporciona cobertura sanitària pública a un total de 141.539 habitants. La diferència en el nombre d'habitants i de persones a les que es dona cobertura sanitària és deguda a que la FSE dona cobertura a les persones dels municipis d'Albons i Bellcaire que no formen part de l'Alt Empordà .

Antecedents

Des de la FSE es vol donar especial rellevància en l'ús de les dades i la presa de decisions basada en evidència científica per tal de millorar la salut poblacional. Des d'aquesta visió i amb la intenció de millorar la recollida d'informació i la transferència d'aquesta a les decisions polítiques comarcals es promou, juntament amb la Universitat de Girona i el GRECS de la Universitat de Girona, la creació d'Indika Pol de Salut i Social de l'Alt Empordà. El Pol de Salut i Social de l'Alt Empordà és un Observatori de Salut i Social que basa la seva estratègia al territori a partir d'un model de treball basat en tres eixos fonamentals: la comunicació, la conversació i l'acció (Cofiño *et al.*, 2012).

L'objectiu principal d'Indika Pol de Salut i Social de l'Alt Empordà és contribuir a la millora de la salut i qualitat de vida de la població, posant especial atenció a la reducció de les desigualtats socials de salut (World Health Organization, WHO, 2010). Per fer-ho cal fer ús dels determinants socials de salut (SDH per les seves sigles en anglès) com a eines per mesurar el progrés en els ODS (Asamblea General de las Naciones Unidas, ONU, 2015).

1.1 DETERMINANTS SOCIALS DE LA SALUT

Els SDH són els factors no mèdics que tenen una influència sobre l'estat de salut. Són les condicions en les quals una persona neix, creix, treballa i viu, així com l'edat i altres factors que tenen un impacte en el dia a dia de les persones. Aquests factors inclouen les polítiques econòmiques, les normes socials, les polítiques socials i els sistemes polítics entre d'altres (WHO, 2020). Diversos estudis demostren com els determinants socials de salut són rellevants en l'augment de l'excés de mortalitat i malalties (Wilkinson i Marmot, 2003). Els SDH engloben un conjunt molt ampli de factors com són les condicions de vida, els sistemes polítics i les estructures econòmiques, culturals i de poder (WHO, 2020).

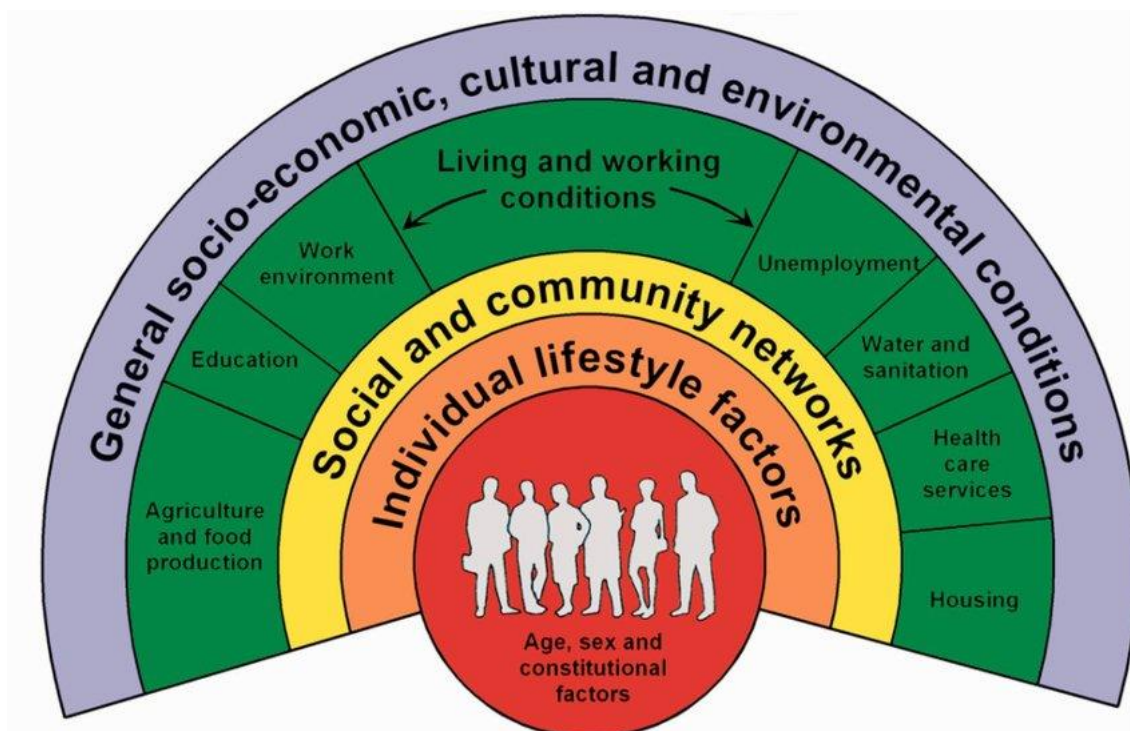


Figura 1 Model de determinants de salut de Dahlgren i Whitehead. Font: Dahlgren i Whitehead (1991).

Dahlgren i Whitehead (1991) proposen un model de determinants de salut estructurat com un arc de sant Martí format per múltiples capes d'influència (Figura 1). Al centre de la figura, els individus posseeixen edat, sexe i característiques constitucionals que influeixen en la seva salut i que són en gran part fixades. No obstant això, al seu voltant hi ha influències teòricament modificables per la política. En primer lloc, hi ha factors de comportament personal, com ara els hàbits de fumar i l'activitat física. En segon lloc, els individus interactuen amb els seus companys i la comunitat immediata i són influenciats per ells, que es representa a la segona capa. A continuació, la capacitat d'una persona per mantenir la seva salut (a la tercera capa) es veu influenciada per les seves condicions de vida i treball, el subministrament d'aliments i l'accés a béns i serveis essencials. Finalment, com a mediador de la salut de la població, prevalen les influències econòmiques, culturals i ambientals en la societat en general. Aquest model per descriure els determinants de la salut fa èmfasi en les interaccions: els estils de vida individuals estan incrustats en les normes i xarxes socials i en les condicions de vida i de treball, que al seu torn estan relacionades amb l'entorn socioeconòmic i cultural més ampli (Dahlgren i Whitehead, 2006).

Antecedents

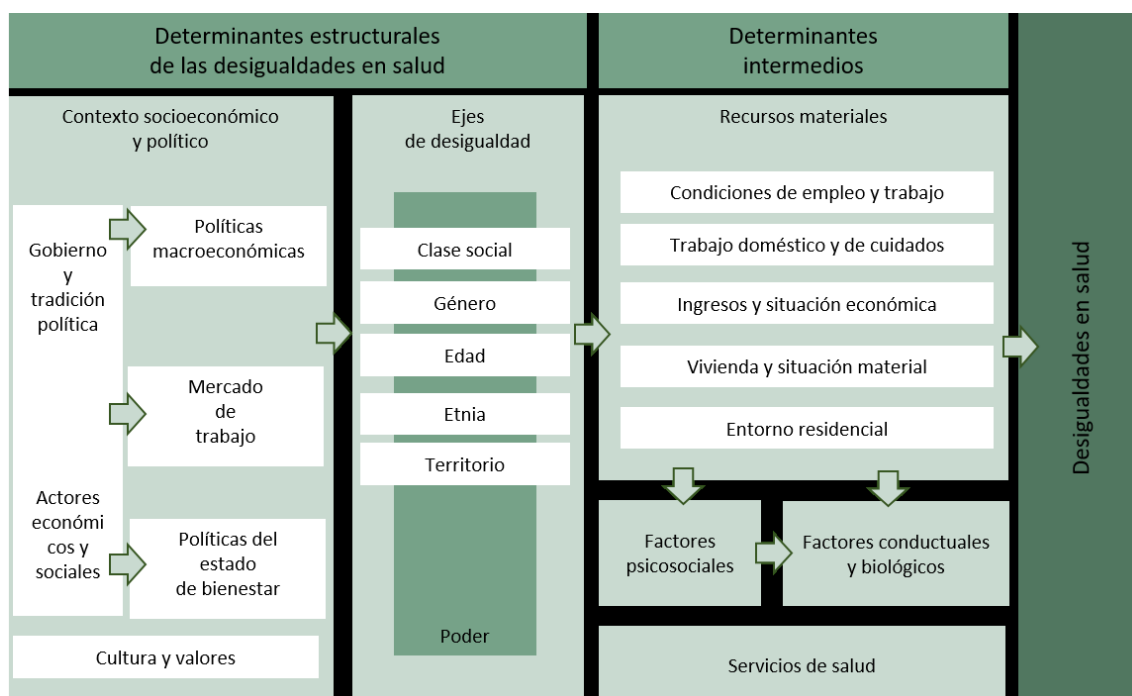


Figura 2 Marc conceptual dels Determinants Socials de la Salut. Font: Comisión para reducir las desigualdades en salud en España (2012).

La “Comisión para reducir las desigualdades en salud en España” (2012) proposa un altre model o marc conceptual sobre el que fonamentar les decisions (Figura 2). Aquest marc conceptual es compon de dos elements principals: els factors estructurals i els factors intermedis de les desigualtats en salut. Els factors estructurals són d’una banda el context socioeconòmic i polític i d’altra banda els eixos de desigualtat. El context socioeconòmic i polític inclou el govern i d’altres factors econòmics que tenen una influència sobre les polítiques generals de l’estat de benestar, economia o mercat de treball entre d’altres. En aquest subapartat també s’inclouen els valors socials i cultures que donen una base filosòfica i moral sobre la qual sustentar les decisions polítiques i les jerarquies de poder de la societat. Els eixos de desigualtat tenen en compte les diferents estructures socials que poden afectar a les oportunitats individuals per tenir una bona salut. Algunes d’aquestes estructures socials són el territori, l’ètnia, l’edat, el gènere o la classe social.

Els determinants intermedis engloben els recursos materials, els factors psicosocials, els factors conductuals i biològics, així com els serveis de salut. Alhora els recursos materials poden subdividir-se en cinc grups que són: (1) les condicions d’ocupació i treball, (2) treball domèstic i de cura, (3) ingressos i

Antecedents

situació econòmica, (4) habitatge i situació material i (5) entorn residencial (Daponte *et al.* 2019).

Les desigualtats socials en salut són diferències sistemàtiques en salut entre diferents grups socioeconòmics d'un país. Sovint aquestes diferències afecten negativament als col·lectius socialment menys afavorits per qüestions socioeconòmiques, d'ètnia o sexe. Donat que no hi ha cap raó biològica per a la seva existència és evident que aquestes estan, en gran manera, configurades per factors estructurals, socioeconòmics o culturals i per tant han de ser adreçades (Whitehead, 2006).

Totes les diferències sistemàtiques en salut entre els diferents grups socioeconòmics d'un país es poden considerar injustes i, per tant, es poden classificar com a desigualtats en salut. No hi ha cap raó biològica per a la seva existència, i és evident que fins i tot les diferències sistemàtiques en els estils de vida entre grups socioeconòmics estan en gran manera configurades per factors estructurals. En resum, les desigualtats socials en salut es generen directament o indirecta per factors socials, econòmics i ambientals i per estils de vida influïts estructuralment. Aquests determinants de les desigualtats socials poden canviar.

La comissió dels determinants socials de la salut va ser creada l'any 2005 per l'OMS amb l'objectiu d'ajudar els països i els associats mundials en pro de la salut a abordar els factors socials que condueixen a la mala salut i les inequitats. La Comissió va assenyalar els determinants socials de la salut com una de les causes més importants de la mala salut i desigualtats en els països i entre ells.

L'any 2008 la comissió va publicar el llibre "*Subsanar las desigualdades en una generación: Alcanzar la equidad sanitaria actuando sobre los determinantes de la salud*" (WHO, 2008). En el llibre, la comissió recomana centrar les accions en tres punts clau: (1) millorar les condicions de vida, (2) lluitar contra la distribució desigual del poder, diners i recursos i (3) mesurar la magnitud del problema amb la intenció d'analitzar i avaluar els efectes de les intervencions realitzades.

Les desigualtats provocades per la forma en com està organitzada la societat fan que les possibilitats de desenvolupar-se a la vida i gaudir d'una bona salut

Antecedents

estiguin mal distribuïdes tant a una mateixa societat com entre les diverses societats. Aquestes desigualtats s'observen en la taxa d'escolarització a la infància, amb la naturalesa de l'ocupació i les condicions de treball a la mitjana edat, així com en les característiques del medi construït i la qualitat del medi natural en el qual viu la població. Segons les característiques d'aquest entorn, les condicions físiques, el suport psicosocial i els esquemes de conducta variaran per a cada grup, fent-los més o menys vulnerables a la malaltia. L'estratificació social també crea disparitats en l'accés al sistema de salut i a la seva utilització, la qual cosa dóna lloc a desigualtats en la promoció de la salut i el benestar, la prevenció de malalties i les possibilitats de restabliment i supervivència després d'una malaltia (WHO, 2009).

La desigualtat de les condicions de vida està determinada per estructures i processos socials més profunds. Les desigualtats són sistemàtiques i són el resultat de normes, polítiques i pràctiques socials que toleren o fins i tot afavoreixen una distribució injusta del poder, la riquesa i altres recursos socials necessaris, i de l'accés a aquests.

L'acció sobre els determinants socials de la salut serà més eficaç si hi ha sistemes de dades bàsiques - com ara els registres civils i programes d'observació sistemàtica de les desigualtats sanitàries i dels determinants socials de la salut - i mecanismes que assegurin que les dades poden ser interpretades i utilitzades en l'elaboració de polítiques, sistemes i programes més eficaços. La sensibilització i la formació en matèria de determinants socials de la salut és essencial.

1.2 OBJECTIUS DE DESENVOLUPAMENT SOSTENIBLE

Els Objectius de Desenvolupament Sostenible (ODS), també coneguts com a Objectius Mundials, es van adoptar per tots els estats membres el 2015 com una crida universal per posar fi a la pobresa, protegir el planeta i garantir que totes les persones gaudeixin de pau i prosperitat per a 2030 (UN Asambleu General de las Naciones Unidas, 2015). Els 17 ODS estan integrats, ja que reconeixen que les intervencions en una àrea afectaran els resultats d'altres i que el desenvolupament ha d'equilibrar la sostenibilitat mediambiental, econòmica i

Antecedents

social. Seguint la promesa de no deixar ningú enrere, els països s'han compromès a accelerar el progrés per a aquells més endarrerits. És per això que els ODS han estat dissenyats per portar al món diversos "zeros" que canviïn la vida, el que inclou pobresa zero, fam zero, SIDA zero i discriminació zero contra les dones i nenes. Tothom és necessari per assolir aquests objectius ambiciosos. Es necessita la creativitat, el coneixement, la tecnologia i els recursos financers de tota la societat per aconseguir els ODS en cada context.

Els 17 ODS són: (1) fi de la pobresa, (2) fam zero, (3) salut i benestar, (4) educació de qualitat, (5) igualtat de gènere, (6) aigua neta i sanejament, (7) energia assequible i no contaminant, (8) treball decent i creixement econòmic, (9) indústria, innovació i infraestructura, (10) reducció de les desigualtats, (11) ciutats i comunitats sostenibles, (12) reproducció i consum responsable, (13) acció pel clima, (14) vida submarina, (15) vida d'ecosistemes terrestres, (16) pau, justícia i institucions sòlides i (17) aliances per assolir els objectius.

En relació amb els ODS, el Grup interinstitucional i d'experts en indicadors d'ODS (IAEG-SDG), va desenvolupar un marc d'indicadors globals aprovats a la 48a sessió de la Comissió d'Estadística de les Nacions Unides, celebrada el març de 2017 (Comisión de Estadística de las Naciones Unidas, 2017).

El marc d'indicadors globals va ser adoptat posteriorment per l'Assemblea General el 6 de juliol de 2017 i es troba a la Resolució adoptada per l'Assemblea General sobre el Treball de la Comissió d'Estadística relativa a l'Agenda 2030 per al Desenvolupament Sostenible (A / RES / 71/313), Annex (ONU, 2020). Segons la Resolució, el marc d'indicadors serà refinat anualment i revisat de manera exhaustiva per la Comissió d'Estadística en la seva cinquanta-primerà sessió del març de 2020 i la seva cinquanta-sisena sessió, que se celebrarà el 2025. El marc d'indicadors global es complementarà amb indicadors a nivell regional i nacional, que desenvoluparan els estats membres.

Els refinaments anuals d'indicadors s'inclouen al marc d'indicadors a mesura que es produeixen. En línia amb el mandat del grup, els IAEG-SDG van proposar 36 canvis importants en el marc en forma de substitucions, revisions, addicions i

Antecedents

suppressions com a part de la revisió integral del 2020, que van ser aprovades per la 51a Comissió d'Estadística el març de 2020 (ONU, 2020).

El marc d'indicadors globals inclou 231 indicadors únics. No obstant això, el nombre total d'indicadors que figuren en el marc global d'indicadors dels indicadors d'ODS és de 247, ja que dotze indicadors es repeteixen sota dos o tres objectius diferents (ONU, 2020).

1.3 OBSERVATORI DE SALUT PÚBLICA

Un observatori de salut pública pot ser entès com una entitat virtual enfocada a la producció d'observacions periòdiques e integrades sobre la població amb l'objectiu de donar suport i guiar les accions efectives per a la millora de la salut pública (Castillo, 2015).

Existeixen diferents tipus d'observatoris de salut pública segons la seva missió, composició, focus o cobertura geogràfica (Jacobson i Castillo, 2014). Les funcions d'un observatori variaran segons el context local en el qual tinguin capacitat d'operar, dels recursos i de les habilitats i capacitats disponibles. Un altre factor important serà el seu estat de desenvolupament (Jacobson i Castillo, 2014). Tot i la possible diversitat en la definició del concepte observatori, hi ha algunes característiques que solen ser comunes entre diferents observatoris com, per exemple, generar instantànies sobre l'estat de salut d'un territori, generar informes sobre les tendències de la població en relació amb SDH concrets, analitzar i interpretar les dades relacionades amb la població objectiu, treballar de forma conjunta amb les entitats i decisors polítics.

A partir del treball de Caiaffa *et al.* (2014), que van realitzar un conjunt de qüestionaris a diversos observatoris de salut, és possible enumerar algunes de les funcions de manera més concreta. Aquestes funcions són: integrar informació local en la presa de decisions sobre salut pública, proporcionar una vista general de l'estat de salut, promoure una cultura basada en accions concretes i basada en dades, proporcionar ajuda a l'hora de solucionar problemes locals, elaborar anàlisis e investigacions científiques i ser una unitat d'intel·ligència de negoci

Antecedents

capaç d'agrupar dades de diferents fonts amb l'objectiu de transformar les dades en coneixement útil.

Per tal de dissenyar Indika s'ha fet una cerca d'observatoris que actuen en diversos nivells administratius, tant nacionals com internacionals. Aquesta cerca s'ha realitzat mitjançant dues fonts d'informació. D'una banda, la literatura científica disponible en diverses revistes indexades a partir dels buscadors GoogleScholar i SciElo. D'altra banda, una cerca a partir dels resultats obtinguts de les paraules clau "*Public Health Observatory*" als buscadors Google i DuckDuckGo. La Taula 1 Llistat d'observatoris mostra els observatoris analitzats.

Taula 1 Llistat d'observatoris analitzats per a la definició d'Indika.

Nom de l'Observatori	Nivell administratiu	URL
Global Health Observatory	Mundial	https://www.who.int/data/gho
Observatorio de Salud de Asturias	Comunitat Autònoma	https://obsaludasturias.com/obsa/
Observatori de Salut i impacte de Polítiques	Municipal (Barcelona)	https://ajuntament.barcelona.cat/observatorisalut/estat-de-salut/indicadors
County Health Rankings i Roadmaps	País (Estats Units d'Amèrica)	https://www.countyhealthrankings.org/
Observatori del Sistema de Salut de Catalunya	Comunitat Autònoma (Catalunya)	http://observatorisalut.gencat.cat/ca/inici

Antecedents

Observatori de les Desigualtats en Salut	Comunitat Autònoma (Catalunya)	http://observatorisalut.gencat.cat/ca/observatori-desigualtats-salut/
Observatori de la Mort	Comunitat Autònoma (Catalunya)	http://observatorisalut.gencat.cat/ca/observatori_mort/
Observatorio de resultados del Servicio Madrileño de Salud	Comunitat Autònoma (Madrid)	https://www.comunidad.madrid/servicios/salud/observatorio-resultados-servicio-madrileno-salud
Observatorio de Salud y Medio Ambiente de Andalucía	Comunitat Autònoma (Andalusia)	https://www.osman.es/
Observatorio de Salud Mental	Comunitat autònoma (Extremadura)	https://www.feafesextremadura.com/observatorio-de-salud-mental/
Observatorio Valenciano de Salud	Comunitat autònoma (Valencia)	https://www.sp.san.gva.es/sscc/opciones4.jsp?CodPunto=3707iOpcion=SANMSOVS1iMenuSup=SANMSiNivel=2
Public Health Wales Observatory	Gal·les (Regne Unit)	http://www.publichealthwalesobservatory.wales.nhs.uk/home

Antecedents

European Observatory on Health Systems and Policies	Europa	https://www.euro.who.int/en/about-us/partners/observatory
Observatorios de Género y Salud en América Latina y el Caribe	Amèrica llatina	https://www.paho.org/es
Observatorio Social "la Caixa"	No Especificat	https://observatoriosociallacaixa.org/
New Zealand Public Health Observatory	Nova Zelanda	www.nzpho.govt.nz

Font: Elaboració pròpia

En total s'han analitzat 16 Observatoris de diferents nacionalitats i nivells administratius, tant nacionals com internacionals.

Des de l'*Observatorio de Salud de Asturias* (OBSA) proposen un model de treball (desenvolupat conjuntament amb els *County Health Ranking* i la Universitat de Wisconsin) sobre el que basar les accions de l'observatori: informar, conversar i actuar (Cofiño *et al.*, 2012).

Antecedents

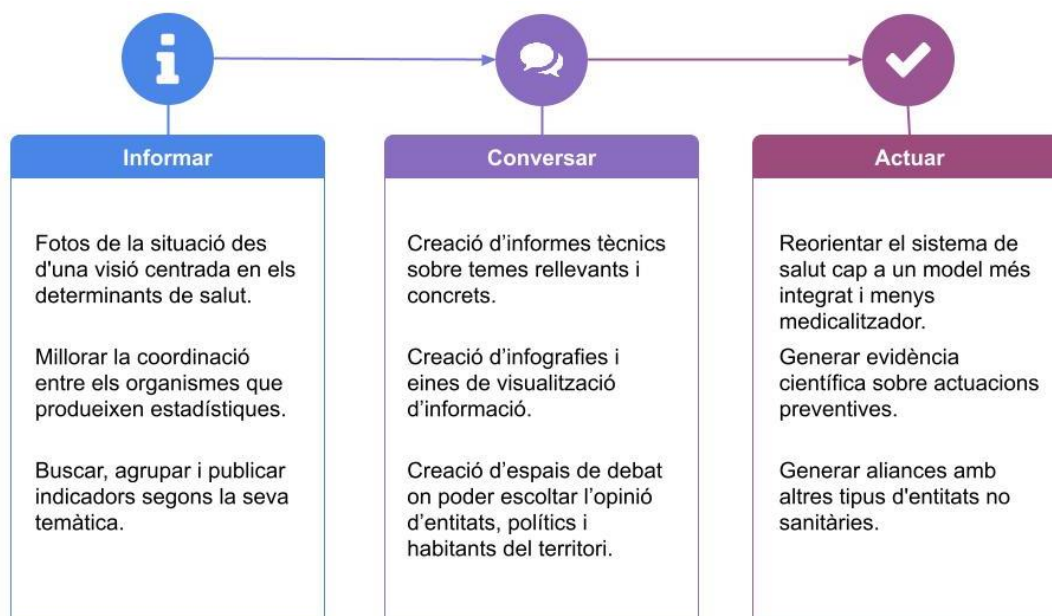


Figura 3 Model de treball proposat per l'OBSA. Font: Elaboració pròpia.

La primera etapa (informar) consisteix a fer saber a la població i als diferents agents del territori, com els polítics o entitats locals, quina és la situació del territori. L'objectiu és iniciar una conversa que faciliti la presa de decisions i ajudi a concretar accions per millorar la salut al territori. Per poder informar correctament a la població és important conèixer els problemes i dinàmiques d'aquesta, així com afegir una perspectiva sobre les riqueses i aspectes positius. No és productiu centrar-se únicament en els aspectes negatius. També és important aprendre dels punts forts de cada territori.

Un cop s'ha informat a la població i actors és necessari iniciar una conversa. Existeixen diverses formes de dur a terme aquesta conversa, però les més comunes solen ser la creació i difusió d'informació en formats que s'adaptin als diferents sectors (com visualitzacions o informes tècnics), la creació d'espais de debat on poden participar tots els agents del territori i la formació en temes de salut comunitària (Cofiño *et al.*, 2012).

L'última etapa és actuar. L'objectiu de les accions ha de ser reorientar el sistema, sobretot el de salut, cap a un model més promotor de la salut, capaç de generar aliances al territori amb la intenció de millorar significativament la salut de forma sostinguda (Cofiño *et al.*, 2012). Donat que encara existeix poca evidència científica respecte a les diverses intervencions, és important desenvolupar una

Antecedents

perspectiva científica d'aquestes actuacions per poder generar l'evidència necessària i que les intervencions siguin extrapolables a altres territoris o comunitats.

Des de l'observatori Indika, l'estratègia per la qual s'ha optat ha estat centrar-se a informar a la població generant informes i articles científics conjuntament amb la Universitat de Girona. Es busca informar mitjançant visualitzacions d'informació similars a les proposades per EUROSTAT (2019, 2020) i Hametner *et al.* (2019).

1.4 REAL WORLD DATA

Es denomina *Real World Data (RWD)* als estudis observacionals basats en dades reals obtingudes de la pràctica clínica diària (Llano *et al.*, 2014) i en la vida diària dels pacients. En concret, els RWD es podrien definir com els estudis que recullen dades rellevants per a la salut humana que no provenen d'assaigs clínics aleatoritzats (RCT) convencionals (The Association of British Pharmaceutical Industry, 2011a; 2011b).

Així, els RWD combinen diferents tipus de fonts de dades: dades obtingudes de la pràctica clínica a partir de cohorts poblacionals (disseny amb dades individuals), dades obtingudes de la pràctica clínica a partir de dades administratives (disseny amb dades individuals), dades de la vida diària dels pacients obtingudes a partir d'enquestes de salut (disseny amb dades individuals) i indicadors contextuais sobre determinants socials i ambientals de la salut (disseny ecològic). La Real World Evidence (RWE) s'entén com l'evidència obtinguda a partir d'aquests estudis.

Els RWD proporcionen la validesa externa de la qual manquen els RCT, els dissenys de major validesa interna. De fet, el principal avantatge dels RWD, a més d'un cost menor, major grandària de la mostra i major representativitat que altres dissenys d'investigació, és la seva alta validesa externa (Staa *et al.*, 2012; Kneale *et al.*, 2016).

Per aconseguir aquesta validesa, tant interna com externa, l'evidència que generarem ha de ser repetible (el que significa que els investigadors han

Antecedents

d'esperar produir resultats idèntics en aplicar la mateixa anàlisi a les mateixes dades per a qualsevol pregunta), reproduïble (un investigador diferent hauria de poder dur a terme la mateixa tasca en l'execució d'una anàlisi donat en una base de dades determinada i esperar produir un resultat idèntic al del primer investigador), replicable (la mateixa pregunta abordada utilitzant l'anàlisi idèntica contra dades similars produeix resultats similars), generalitzable (circumstància en la qual s'elaboren anàlisis idèntiques en diferents bases de dades i es mostren resultats consistentment similars), robusta (el que vol dir que els resultats no han de ser massa sensibles a les eleccions subjectives que es poden fer dins d'una anàlisi) i calibrada (un estudi observacional sempre ha d'anar acompanyat de diagnòstics que provin els supòsits al voltant del disseny, els mètodes i les dades).

Hi ha una sèrie de problemes que impedeixen aprofitar el veritable potencial dels RWD, entre els quals caldria citar la manca d'un procediment estandarditzat de recollida de dades, la manca d'estàndards de qualitat de les dades i dels processos de validació d'aquests, la manca de bases de dades representatives i la manca de suficients estudis que mostrin com es poden utilitzar els RWD en els sistemes de salut depenent del seu objectiu (per exemple, per a la presa de decisions clíniques, presa de decisions reguladores, avaluació de tecnologies de la salut, etc.). No obstant això, si s'aborden aquests desafiaments, els RWD poden proporcionar nous coneixements sobre els patrons de malalties i ajudar a millorar la seguretat i l'eficàcia de les intervencions mèdiques.

És evident que els RWD es corresponen amb l'activitat investigadora de qualsevol estadístic aplicat. En aquest sentit, els RWD combinen dades procedents de diverses fonts, els estandarditzen en un model comú de dades i permeten elaborar anàlisis sistemàtiques de forma estàndard.

Tot i que els RWD sempre han tingut una gran importància, és en els últims anys, gràcies al progrés de l'electrònica i de la informàtica que han permès no només processar sinó també recollir un gran volum de dades clíniques que han cobrat un gran renom. Durant l'última dècada s'han introduït amb força nous sistemes d'informació sanitària que han suposat un increment notable tant en la qualitat com en la quantitat de les dades clíniques informatitzades. La digitalització i

Antecedents

integració de la informació (històries clíniques, registres de pacients, dades d'assegurances mèdiques, enquestes de salut, dades administratives, biobancs, salut mòbil, etc.) recolzada en la Intel·ligència Artificial, ha facilitat l'obtenció de dades en un context real (RWD), que analitzats de forma adequada poden generar evidència científica (RWE).

La gran sofisticació i volum de dades amb què comptem actualment ha fet necessari treballar amb equips multidisciplinaris per poder treure el màxim profit i el millor coneixement possible de les dades. Així, en la presa de decisions, a diferència d'abans on la participació era escassa, han passat a cobrar importància els economistes de la salut, els estadístics, els matemàtics, els biomèdics, els gestors, els polítics, els enginyers, els pacients, els observadors i els cuidadors, entre d'altres.

Atès que les dades procedeixen de diverses fonts, combinar-les representa un esforç d'unificació de formats, de codificació dels continguts, així com d'homogeneïtzació de les organitzacions lògiques i terminologies utilitzades per descriure les variables. Més complicat resulta quan entre les fonts de RWD es troben fonts de dades externes (dades generades per aplicacions mòbils, *wearables*, dades procedents d'enquestes de salut poblacionals, variables contextuais procedents de censos de població, etc.). Tampoc resulta fàcil quan es pretenen utilitzar els RWD de diferents centres que poden procedir de diferents comunitats autònomes, països, etc., cada un amb un sistema d'informació diferent.

Caldria un model comú de dades (CDM per les seves sigles en anglés) que permetés l'anàlisi sistemàtica de bases de dades d'estudis observacionals disperss. Actualment, existeixen diverses iniciatives relacionades amb aquest concepte com són, per exemple, OpenEHR (Kalra *et al.*, 2005) o OHDSI (Hripcsak *et al.*, 2015). Els objectius d'un CDM són, entre d'altres: transformar les dades contingudes en les diferents bases de dades en un format comú (model de dades); proporcionar una representació comuna dels mateixos (terminologies, vocabularis, esquemes de codificació); i realitzar anàlisis sistemàtiques utilitzant una biblioteca de rutines analítiques estàndard escrita basant-se en el format comú.

Antecedents

Els beneficis del CDM rau en el seu ús per crear repositoris de dades que facin servir un únic sistema de metadades, així com transformar les dades existents en l'esquema estàndard del CDM. L'eficiència que s'obté de l'estandardització pot accelerar i agilitzar tot el que es faci a continuació amb les dades. A més, al permetre generar evidència d'una àmplia varietat de fonts, afavoreix la investigació col·laborativa.

Per acabar, és important dir que en els RWD ens podem trobar diferents tipus de biaixos tal com el biaix de selecció, el biaix de confusió, el biaix d'informació (també conegut com de classificació), etc.

En els RWD, com a tots els dissenys observacionals, és molt probable que es produeixi un biaix de selecció. Alguns subjectes tenen més probabilitat de ser observats i, per tant, d'estar presents en la mostra que altres. El biaix de selecció és un error sistemàtic que es dona quan la distribució de freqüències de les variables explicatives és sistemàticament diferent, bé entre els individus, bé entre grups (en particular entre els grups de tractament i control). Atès que el resultat d'interès és el resultat de l'efecte del tractament després de controlar les diferències en les variables explicatives, és possible que una comparació simple de resultats entre individus o grups no proporcioni una estimació no esbiaixada de l'efecte del tractament quan les variables explicatives es distribueixen de diferent manera entre aquests individus o grups.

Si la selecció fos exògena, és a dir, si la probabilitat que s'observés a un subjecte fos idèntica per a tots els subjectes, n'hi hauria prou amb ponderar la mostra de tal manera que s'atorgués menor pes als subjectes efectivament observats. No obstant això, és molt probable que els subjectes observats ho siguin a causa de factors no observats que influeixin precisament en aquesta observació i que estarien correlacionats amb els factors no observables que afectessin la variable resposta. En tot cas, la probabilitat que aquests subjectes hagin estat observats no és la mateixa per a tots ells. Així doncs, la ponderació (estandardització) per edat i sexe no corregiria el biaix de selecció. En aquest cas (denominat selecció endògena), s'han d'utilitzar mètodes estadístics adequats més complexos (models en dues parts, etc.).

Antecedents

El biaix de selecció és l'origen del biaix de confusió. Hi ha biaix de confusió quan observem una associació no causal entre l'exposició i l'esdeveniment en estudi o quan no vam observar una associació real entre l'exposició i l'esdeveniment en estudi per l'acció d'una tercera variable que no és controlada (anomenada confusor o factor de confusió).

Els errors que poden ocórrer en el mesurament, tant en l'exposició com en el resultat, ens conduirien al biaix d'informació o de classificació. Els errors als quals fem referència consistirien, d'una banda, en identificar els subjectes que no van experimentar el resultat com si ho haguessin fet (o estar exposat quan no ho estan) i, d'altra banda, en suposar que alguns subjectes no han experimentat el resultat quan en realitat ho van experimentar (o no estar exposats quan ho estan). El biaix d'informació pot ser no diferencial (aleatori) quan el grau d'error és el mateix per als grups que es comparen; i diferencial (no aleatori) quan el grau d'error és diferent per als grups estudiats. El biaix de classificació es refereix particularment a aquest últim tipus.

Hi ha diferents mètodes per minimitzar aquests biaixos. Entre els diferents mètodes per controlar el biaix de selecció, destacarem aquells classificats en mètodes de disseny i en mètodes d'anàlisi, els quals podríem classificar en mètodes basats en observables i en mètodes basats en no observables. Òbviament, tots aquests mètodes solen combinar-se. Pel que fa al biaix de classificació, els principals mètodes proposats per a la seva minimització consisteixen a establir procediments estandarditzats de recollida de dades; implementar mesures/marcs estàndard per al mesurament (per exemple, utilització de classificacions com CIE-10, ATP, etc.), etc.

1.5 MODELS DE DADES BASATS EN ESDEVENIMENTS

Els RWD s'obtenen a partir de múltiples fonts de dades. De forma genèrica, podem diferenciar entre aquelles fonts de dades pròpies dels entorns mèdics com: registres de salut electrònics, dades clíniques (analítiques, etc.) i dades de facturació. O bé, fonts externes als registres mèdics habituals com: registres de mortalitat, fonts de dades obertes (Demski, 2016), dades que provenen de telèfons intel·ligents, *wearables* (per exemple rellotges intel·ligents) i altres

Antecedents

dispositius mòbils (Reeder i David, 2016) i dades procedents de sensors que conformen l'anomenada Internet of Things (Scarpato et al., 2017). Entre les fonts de dades obertes es troben les bases de dades obertes sobre la qualitat de l'aire (Hasenkopf et al., 2015) o temperatura, entre d'altres.

Donada la definició anterior es dedueixen algunes característiques dels RWD. En ésser recopilades de diferents fonts el volum de dades pot ser molt elevat, sobretot si les fonts són dispositius mòbils. Recopilar informació de diferents fonts de dades implica la necessitat de gestionar múltiples formats (JSON, XML, CSV, entre d'altres). Aquesta definició encaixa amb el concepte de dades complexes, definit com un conjunt d'informació que no és únicament numèrica o simbòlica, sinó multimèdia i multiestructura. L'emmagatzematge i tractament d'aquest tipus d'informació requereix estructures adequades (Boukraâ i Boussaid, 2009).

En epidemiologia, un estudi de cohorts és un tipus de disseny observacional i analític (prospectiu o retrospectiu). El que defineix els estudis de cohort és que els subjectes de l'estudi es trien segons l'exposició d'interès. En el cas més senzill, se selecciona un grup exposat i un grup no exposat i se segueixen ambdós al llarg del temps per comparar l'ocurrència d'un esdeveniment d'interès. Des d'una perspectiva de l'ús d'una arquitectura basada en esdeveniments podem pensar en un estudi de cohorts com una entitat que reacciona a diferents esdeveniments produïts al món exterior (Fowler, 2006). Això permet actualitzar una cohort a partir de notificacions de fonts externes.

És, en aquest context, en el que es proposa la biblioteca de codi obert eventr (Franquet *et al.*, 2020) de R (R Core Team, 2020) que facilita la implementació d'arquitectures dirigides per esdeveniments (EDA per les seves sigles en anglès).

En estructurar la informació mitjançant un model EDA s'obtenen els següents avantatges: s'obté un registre d'auditoria (*audit trail*), simplicitat a l'hora de depurar el codi, la capacitat de crear estats alternatius i l'ús d'esdeveniments retroactius (Fowler, 2017; Erb *et al.*, 2016).

Antecedents

- L'*audit trail* és un registre de tots els esdeveniments que han tingut lloc sobre un objecte d'interès. Els esdeveniments registren quins canvis s'han fet i en quin moment han tingut lloc.
- En tractar-se d'esdeveniments independents entre si, és més senzill depurar errors de codi. Ofereix la possibilitat de crear entorns de prova on es reproduïxin exactament tots els canvis que han tingut lloc i retrocedint o avançant fins al moment en què hagin ocorregut els errors.
- Utilitzar esdeveniments de forma retroactiva permet tornar a calcular l'estat d'un objecte en un moment anterior del temps de forma més senzilla. Això és especialment útil en els estudis de cohorts retrospectius.
- Les arquitectures dirigides per esdeveniments permeten la creació d'estats alternatius. Podem modificar la definició d'un dels esdeveniments i tornar a calcular l'estat de l'objecte d'interès de forma retroactiva.

Tenint en compte les qualitats d'aquest tipus d'arquitectura, és més senzilla la creació i anàlisi de dades a partir de RWD. Donada la complexitat i el cost dels estudis on s'utilitzen aquest tipus de dades, són necessaris per tenir la capacitat de tornar a computar els estats dels individus en cas d'existir errors de codi, així com tenir la capacitat d'auditar tots els passos que s'han portat a cap per assegurar la reproductibilitat d'aquests estudis. Per altra part, la creació d'estats alternatius facilita la simulació de pacients sintètics.

En la secció 3.4 "Implementació d'una llibreria de R per al modelatge de dades basades en esdeveniments" es descriu amb més detall la implementació realitzada i en l'apartat 4.2 s'explica com fer ús d'aquesta llibreria per tal de modelar les dades fent ús de la mateixa en un cas simulat i un de real.

1.6 VISUALITZACIÓ D'INFORMACIÓ

En aquest apartat es descriu que és una visualització i quines característiques ha de tenir aquesta. En aquest text s'utilitzaran de forma indistinta les paraules visualització i infografia per referir-se al mateix terme.

Existeixen diverses definicions d'infografia, però al llarg d'aquest text la definició que es farà servir és la proposada per Cairo (2016): "Una visualització és

Antecedents

qualsevol representació visual d'informació dissenyada per comunicar, analitzar, descobrir i explorar les dades". Aquesta definició inclou el concepte d'infografia tradicional que podem trobar en diaris o revistes i les visualitzacions realitzades per a altres mitjans com poden ser els navegadors web o telèfons mòbils intel·ligents.

1.6.1 Característiques d'una visualització

Per tal de complir amb el seu propòsit una infografia ha de ser: veraç, funcional, atractiva, astuta i aclaridora.

Una visualització ha de ser veraç. Això vol dir que les dades que s'hi presenten i la forma en què es presenten ha de ser el més ajustat a la realitat possible. Una visualització feta amb dades inventades o que no representen la realitat no representa informació i per tant no pot ser considerada com a infografia.

Donat que tota visualització ha de tenir un objectiu és clau que compleixi aquest objectiu. Per tant una visualització ha de ser funcional, és a dir, ha de complir amb el seu objectiu. Si volem explicar què és el canvi climàtic i quins són els seus efectes, la visualització ha d'aconseguir que els usuaris finals tinguin la capacitat, gràcies a la infografia, d'entendre aquest fenomen.

Una visualització ha de ser visualment atractiva. Una de les funcions de tota visualització és, a més a més d'informar o mostrar una informació; de cridar l'atenció de l'usuari. Cal provocar la curiositat de les persones que es troben amb la visualització per tal que tinguin ganes de llegir-la.

Un dels avantatges de fer ús de gràfics per representar la informació és la seva capacitat de mostrar patrons i característiques de forma simplificada. A aquesta qualitat li direm ser astuta.

Si es compleixen les quatre característiques anteriors, una bona visualització serà, a més a més, aclaridora. Això vol dir que, igual que un llibre tècnic, una infografia ens ha de permetre al seu públic objectiu descobrir alguna cosa que no sabien de manera que pugui canviar la seva forma de veure el món amb relació a un tema.

Antecedents

1.6.2 Metodologia en el procés de creació d'una infografia

A continuació es descriu la metodologia de treball que s'ha fet servir per al desenvolupament de les infografies. Per dur a terme les infografies s'ha fet servir la metodologia proposada per la Generalitat de Catalunya a la "Guia de visualització de dades" (Generalitat de Catalunya, 2018). Aquesta metodologia proposa dividir la feina en tres fases que alhora es componen de diverses tasques. Aquestes fases són: la definició de l'estratègia, la recopilació i transformació de les dades i el disseny de la visualització.

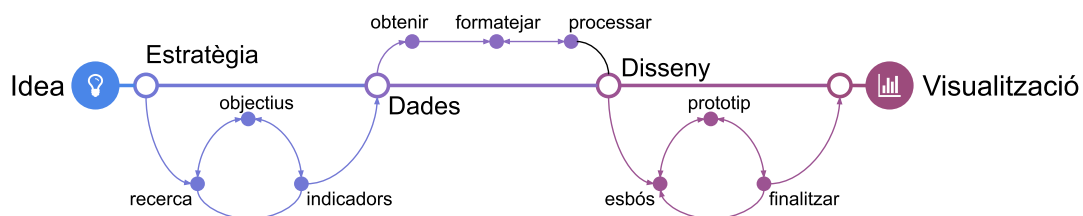


Figura 4 Model del procés de creació d'una infografia. Font: Elaboració pròpia.

1.6.2.1 Estratègia

L'estratègia consisteix a definir la temàtica i quins són els objectius de la visualització. Per fer això cal realitzar una recerca bibliogràfica sobre el tema a tractar. La recerca bibliogràfica ha de servir per definir l'abast de la infografia, ens ajuda a plantejar les preguntes adequades i ens proporciona respostes a aquestes preguntes. Sempre que sigui possible és recomanable parlar amb experts sobre la temàtica a desenvolupar.

Amb la recerca bibliogràfica i la consulta amb els experts s'han de poder definir les preguntes a les quals volem donar resposta. Amb aquestes preguntes de partida se seleccionen els indicadors que ens permetran respondre-les.

La "Guia de visualització de dades" proposa fer ús de l'esquema 5W1H. Aquesta eina pròpia del periodisme consisteix en respondre a les preguntes: Que ha passat? Qui n'és el responsable? On ha passat? Quan ha passat? Perquè quina raó ha passat? Com ha passat?

Antecedents

1.6.2.2 Dades

Un cop definit l'objectiu i concretat a partir d'un conjunt de preguntes a respondre i els seus indicadors corresponents, cal fer: la recerca, transformació i explotació de les dades.

L'obtenció de les dades es pot realitzar a partir de fonts de dades públiques o d'altres fonts d'informació com poden ser bases de dades d'una o diverses entitats. La infografia “La demografia a l'Alt Empordà” explicada amb més detall a la secció 4.1.1.1 és un exemple d'infografia creada a partir de dades públiques, en aquest cas d'Idescat. La infografia “La Natalitat a l'Alt Empordà” explicada a la secció 4.1.1.2 i disponible a l'Annex Infografia “Natalitat a l'Alt Empordà” o bé a la pàgina web d'Indika és un exemple on s'ha fet ús de dades públiques proporcionades per Idescat i dades pròpies de la Fundació Salut Empordà.

Els formats més habituals són CSV, JSON o XML. Per tal de fer-ne l'explotació caldrà transformar aquestes dades a una estructura més adequada segons les nostres necessitats. En el que respecta a les infografies realitzades en aquesta tesis la transformació i explotació de les dades s'ha realitzat mitjançant el llenguatge de programació R (R Core Team, 2020) i el conjunt de llibreries tidyverse (Wickham *et al.*, 2019).

1.6.2.3 Disseny

El disseny és la part final de procés. Consisteix a crear la visualització pròpiament dita. Per realitzar aquest pas s'ha fet ús d'una metodologia iterativa que consisteix a esbossar, prototipar i finalitzar.

Esbossar és un procés que té com a objectiu permetre al dissenyador descobrir quina forma volem donar a la visualització, quins gràfics volem fer servir i on s'ha de posar cada gràfic per tal que la visualització tingui coherència. Un esbós pot

Antecedents

ser un dibuix en un paper o qualsevol altra eina que permeti plasmar la idea en algun mitjà fàcil de compartir i que no impliqui dedicar molt de temps.

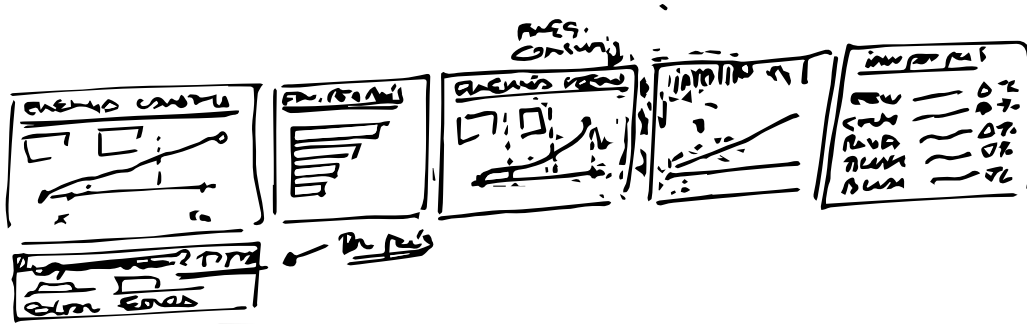


Figura 5 Esbós d'una infografia. Font: Elaboració pròpia.

Un cop s'ha realitzat l'esbós i hi ha una idea general de com ha de ser la visualització es realitza un prototip mitjançant les eines gràfiques o informàtiques adequades. Per la realització de les infografies presentades en aquesta tesi s'ha fet servir una combinació de R (R Core Team, 2020) i el paquet ggplot2 (Wickham, 2016) per realitzar les gràfiques en format SVG. Un cop obtinguts els gràfics la maquetació s'ha realitzat fent ús del programari Adobe Illustrator (Adobe Inc., 2019) i *Inkscape* (Inkscape Project, 2020). El prototip té com a objectiu ser compartit amb els experts implicats en el procés de creació.

Finalment cal revisar els textos i confirmar que la visualització és funcional. En aquesta fase és recomanable buscar persones amb un perfil similar al del públic objectiu per tal de confirmar que la seva comprensió és l'esperada. La "Guia de visualització de dades" de la Generalitat de Catalunya recomana un total de cinc persones per cada iteració, ja que d'aquesta manera és possible trobar els canvis més rellevants a desenvolupar sense la necessitat de fer un ús excessiu de recursos.

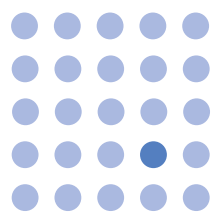
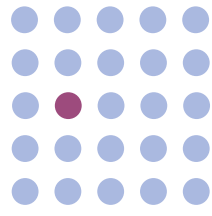
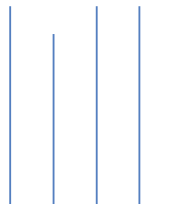
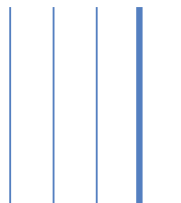
1.6.3 Principis de la visualització d'informació

En aquest apartat es descriuen alguns dels principis fonamentals en la visualització de la informació des d'un punt de vista neurològic. L'objectiu és

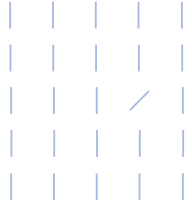
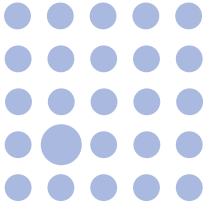
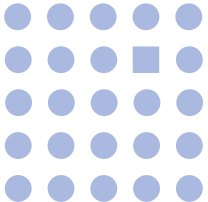
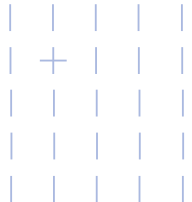
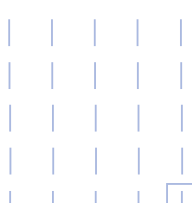

Antecedents

explicar com el nostre cervell interpreta la informació que rep a través del sentit de la vista. Few (2013) fa una selecció d'onze atributs de la percepció visual agrupats en quatre grans grups: color, forma, posició i moviment.

Taula 2 Atributs d'atenció prèvia de la percepció visual.

Categoria	Atribut	Il·lustració
Color	Tonalitat	
	Intensitat	
Forma	Alçada de la línia	
	Amplada de la línia	

Antecedents

	Orientació	
	Mida	
	Figura	
	Marques	
	Encerclament	
Posició	Localització 2D	

Antecedents

Moviment	Parpelleig	El cursor del ratolí en els editors de text modern parpelleja per tal de mostrar en quina part del text ens trobem.
-----------------	------------	---

Font: Elaboració pròpia a partir de Few (2013).

1.7 OPEN DATA

La *Open Knowledge Foundation* (OKF) defineix les dades obertes com aquelles dades que poden ser utilitzades, reutilitzades i redistribuïdes lliurement per qualsevol persona sempre que es doni el crèdit adequat i es distribueixin de la mateixa forma en què han estat publicades (Open Knowledge Foundation, 2021a).

Per considerar que unes dades són “obertes” cal que s’ajustin a tres criteris: (1) El criteri de disponibilitat i accés; (2) el criteri de reutilització i redistribució; i (3) el criteri de participació universal.

El criteri de disponibilitat i accés garanteix que la informació està disponible com un tot i a un cost raonable de reproducció, preferiblement si es pot descarregar a través d’internet. A més a més la informació ha d’estar en un format convenient per a la seva utilització.

El criteri de reutilització i redistribució garanteix que les dades permeten la seva reutilització i redistribució. Inclús és possible integrar-los amb altres conjunts de dades.

Finalment, el criteri de participació universal garanteix que tothom pot fer ús de les dades sense cap mena de discriminació a grups o persones. Les dades han de tenir restriccions a la comercialització de les mateixes per evitar que d’altres

Antecedents

persones paguin per un recurs que és obert. Cal especificar quin ús es pot fer de les mateixes fent ús d'alguna llicència específica.

Les dades obertes han de permetre la interoperabilitat amb altres dades, entitats, persones o agents. La interoperabilitat és la capacitat d'integrar diferents conjunts de dades. La interoperabilitat, en l'àmbit de les dades obertes, és la capacitat d'aquestes per ser combinades amb altres dades i desenvolupar millors productes i serveis.

Les dades obertes han de ser de caràcter no personal. No han de contenir informació sobre individus específics. Altres tipus de dades han d'estar protegides o tenir restriccions especials, com per exemple aquelles que siguin importants per a la seguretat nacional.

En el procés d'obertura de dades per part d'una institució no és necessari que les dades s'obrin totes de cop. És millor començar obrint un conjunt de dades petit i fàcil d'aconseguir.

Un cop obertes les dades és important relacionar-se amb els usuaris i persones que en faran ús. Serà necessari entendre quines són les necessitats dels usuaris amb l'objectiu de descobrir quin altre conjunt de dades els pot ser d'utilitat.

Cal tenir present que gran part de les dades que es publiquen no arriben als usuaris de forma directa, sinó a través d'agents intermedis que transformaran aquestes dades en altres productes com poden ser infografies, pàgines web o altre tipus de suports.

L'OKF defineix en la seva d'obertura de dades quatre fases: (1) Fer una tria de les dades a obrir, (2) seleccionar una llicència oberta que delimiti quins són els drets de propietat intel·lectual de la informació, (3) transformar les dades a un format fàcilment transferible i explotable com per exemple JSON o XML i (4) fer-la visible, publicar a la web de l'entitat, a les xarxes socials o en un catàleg central de bases de dades.

Segons la magnitud de la institució que té interès a oferir dades, realitzar una tria de les dades a obrir pot ser un procés molt costós a causa de la mateixa

Antecedents

complexitat de les dades emmagatzemades. En aquests casos el més recomanable és fer un llistat de dades disponibles i posteriorment relacionar-la amb una llista creada a partir d'una consulta a la comunitat i les persones que en faran ús i seleccionar aquelles que més s'ajusten a la demanda.

A l'hora de publicar la informació és important decidir amb quin tipus de llicència es farà. La Taula 3 mostra un llistat de llicències disponibles que poden ser utilitzades a l'hora d'obrir conjunts de dades i que compleixen la definició d'"obertes" segons l'OKF.

Taula 3 Llicències aptes per a dades obertes.

Nom de la llicència	Requereix atribució	Requereixen compartir igual
Creative Commons CCZero (CC0-1.0)	No	No
Open Data Commons Public Domain Dedication and Licence (PDDL-1.0)	No	No
Creative Commons Attribution 4.0 (CC-BY-4.0)	Si	No
Open Data Commons Attribution License (ODC-By-1.0)	Si	No
Creative Commons Attribution Share-Alike 4.0 (CC-BY-SA-4.0)	Si	Si
Open Data Commons Open Database License (ODbL-1.0)	Si	Si

Font: Elaboració pròpia a partir de Open Knowledge Foundation - Commformant Licenses

Per poder considerar que les dades són obertes cal que aquestes estiguin obertes legalment i tècnicament. Que les dades estiguin obertes tècnicament

Antecedents

significa que aquestes estiguin disponibles en un format intel·ligible per una màquina.

1.7.1 Sistema de classificació de dades obertes

Tim Berners-Lee descriu un esquema de desenvolupament de dades obertes basat en la classificació de les dades mitjançant un sistema d'estrelles. La Taula 4 mostra la classificació per estrelles descrita a la pàgina web 5stardata (<https://5stardata.info/en/>). La columna "Nombre d'estrelles" indica el nombre d'estrelles que s'assignaria a un conjunt de dades que compleix amb els requeriments definits a la columna "Descripció".

Taula 4 Classificació de dades obertes mitjançant el sistema d'estrelles.

Nombre d'estrelles	Descripció
1	Les dades estan disponibles a la web (en qualsevol format) sota una llicència oberta.
2	Les dades estan disponibles de forma estructurada (per exemple, en format Excel en lloc d'en forma d'imatge).
3	Les dades estan disponibles en un format obert no propietari (per exemple, CSV en lloc d'Excel).
4	Utilitzeu un <i>Identificador de Recursos Uniforme</i> (URI per les seves sigles en anglès) de manera que la gent pugui referenciar les dades.

Antecedents

5	Enllaceu les vostres dades amb altres dades per proporcionar context (com per exemple RDFa).
---	--

Font: <https://5stardata.info/en/>

1.7.2 La Notació d'Objectes JavaScript

La Notació d'objectes JavaScript (JSON) (ECMA International, 2017) és una sintaxi lleugera, basada en text, independent del llenguatge per definir formats d'intercanvi de dades. Va ser derivat del llenguatge de programació ECMAScript (ECMA International, 2017), però és independent del llenguatge de programació. JSON defineix un petit conjunt de regles d'estructuració per a la representació portàtil de dades estructurades.

La sintaxi JSON es basa en dues estructures bàsiques: (1) Una col·lecció de parells nombre/valor i (2) una llista ordenada de valors.

Els objectes es defineixen mitjançant els claudàtors tal com mostra la Figura 6.

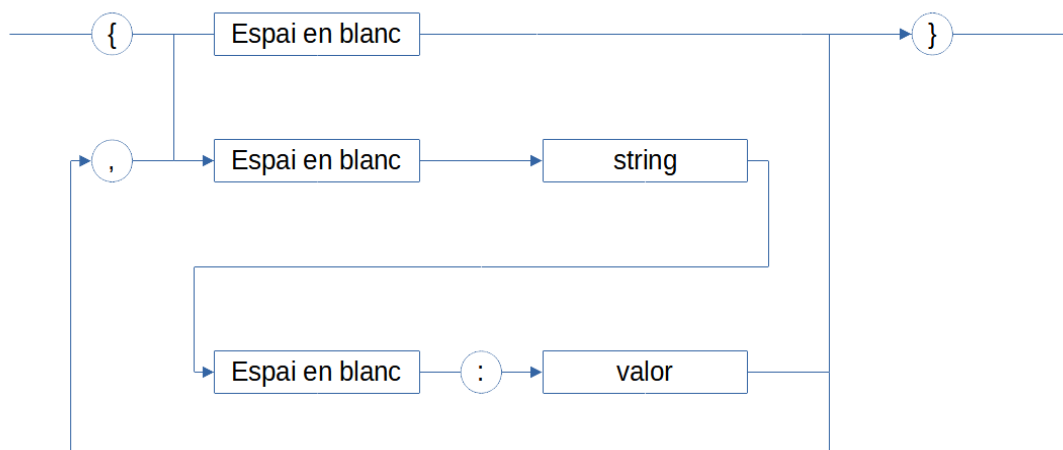


Figura 6 Esquema d'objecte JSON. Font: Elaboració pròpia a partir de <https://www.json.org/json-en.html>

```
{
  "firstName" : "Alvaro",
  "secondName" : "Franquet"
}
```

Antecedents



Una matriu és una col·lecció ordenada de valors. Una matriu comença amb el claudàtor esquerre ([]) i acaba amb un claudàtor dret (]). Entre claudàtors hi ha els nombres separats per comes. La Figura 7 mostra la representació d'una matriu.

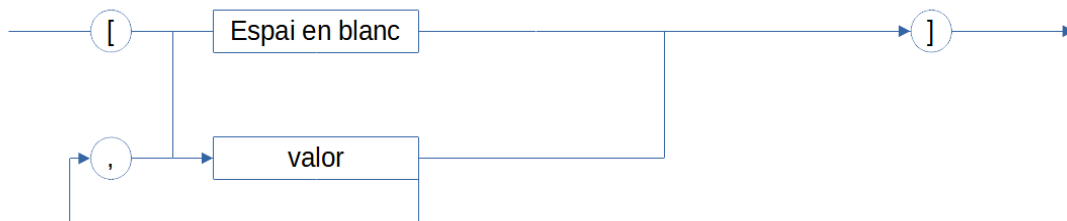


Figura 7 Esquema de definició d'una matriu mitjançant JSON. Font: Elaboració pròpia a partir de <https://www.json.org/json-en.html>



Un valor pot ser de tipus: *string* entre cometes (""), un nombre, un valor booleà (*true* o *false*), un valor nul, un objecte o una matriu. Aquestes estructures es poden imbricar.

1.7.3 El Llenguatge de marcatge extensible

El llenguatge de marcatge extensible (XML) és una sintaxi definida pel *World Wide Web Consortium* (W3C) (Extensible Markup Language (XML) 1.0 (Fifth Edition), 2008). El XML és un llenguatge dissenyat per emmagatzemar i transferir dades fent ús de la xarxa amb la intenció que sigui llegible tant per humans com per màquines.

Igual que el llenguatge HTML, XML fa ús d'etiquetes per definir les estructures contingudes en el document, que a diferència del primer no estan definides

Antecedents

prèviament, és a dir, és l'autor del document XML qui s'encarrega de definir les seves pròpies etiquetes.

Una etiqueta és una cadena de caràcters que està continguda entre els símbols menor que (<) i major que (>). La Figura 8 mostra l'esquema relatiu a la definició d'aquestes etiquetes.

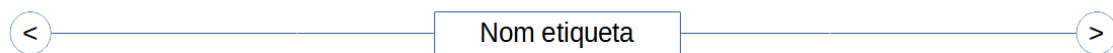


Figura 8 Esquema d'una etiqueta en format XML. Font: Elaboració pròpia

```
<firstName>
```

Un element és un conjunt format per dues etiquetes. Una etiqueta que indica en quin moment comença l'element i una etiqueta que indica en quin moment acaba l'element caracteritzada per l'ús de la barra (/) seguida del nom de l'etiqueta. La Figura 9 mostra l'esquema general per a la creació d'un element.

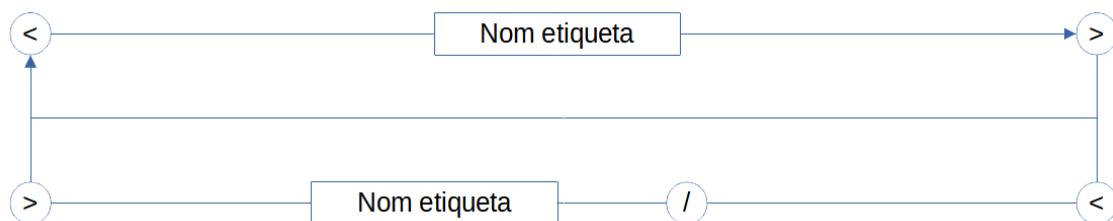


Figura 9 Esquema d'un element en format XML. Font: Elaboració pròpia.

```
<person>  
  <firstName> Álvaro </firstName>  
  <secondName> Franquet </secondName>
```

Antecedents

```
</person >
```

Els elements poden contenir atributs. Els atributs han d'anar entre cometes. Tant els elements com els atributs poden contenir la mateixa informació. Per aquesta raó no hi ha regles clares sobre l'ús d'un o l'altre en cada situació. Sí que hi ha algunes característiques que poden ajudar a l'autor d'un document a decidir sobre l'ús d'un o de l'altre: (1) els atributs només poden tenir un valor mentre que els elements en poden tenir més d'un, (2) els atributs no poden imbricar-se mentre que els elements sí i (3) els atributs són més difícils d'expandir. La Figura 10 mostra l'esquema bàsic d'un element amb atributs.

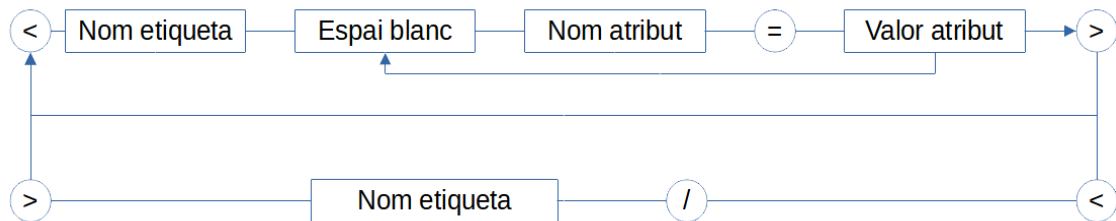


Figura 10 Esquema d'un element en format XML amb atributs. Font: Elaboració pròpia.

```
<person sex = "male">  
  <firstName> Álvaro </firstName>  
  <secondName> Franquet </secondName>  
</person >
```


2 HIPÒTESIS I OBJECTIUS

2.1 HIPÒTESIS

Hipòtesis principals

- És possible la integració de dades associades a l'observatori de salut i *real world data*.
- L'explotació d'aquesta informació integrada permetrà millorar l'estat de benestar de la comarca.

Hipòtesis secundaries

- L'aplicació de mètodes estadístics adients permetrà millorar la qualitat i evitarà biaixos, la qual cosa permetrà aconseguir la validesa interna de la base de dades.
- Mitjançant la reproductibilitat s'aconseguirà la validesa externa d'aquesta base de dades.

2.2 OBJECTIUS

El pol de salut Indika neix amb el propòsit d'afavorir la generació de coneixement de l'impacte en la salut dels determinants socials. Vol ser una ajuda a l'hora de formular polítiques de salut efectives mitjançant l'ús d'indicadors de salut com a eina vehicular del projecte. L'ús d'aquests indicadors ha de permetre orientar estratègies intersectorials i alinear els agents de salut i socials per treballar de manera coordinada.

Objectius generals

- Obtenir una unitat de caràcter científic-tècnic, observatori de salut, que promourà l'anàlisi de la situació de salut i els seus factors determinants, especialment els socioeconòmics, a l'Alt Empordà.

Hipòtesis i objectius

- Subministrar la informació provinent de la integració de l'observatori de salut i del *Real World Data* a entitats públiques i privades que puguin influir per millorar l'estat de benestar de la comarca de l'Alt Empordà.

Objectius específics

- Construir i mantenir un repositori d'informació sanitària, socioeconòmica i mediambiental ubicat a un lloc centralitzat que permeti guardar, organitzar, mantenir i difondre informació digital de l'àmbit sanitari, socioeconòmic i mediambiental de l'Alt Empordà fent referència també a Catalunya, Espanya i Europa.
- Dissenyar una base d'indicadors de salut, socioeconòmics i mediambientals composta per:
 - Resultats de salut: mortalitat, morbiditat (objectiva i auto percebuda).
 - Determinants de la salut: qualitat assistencial, estils de vida.
 - Factors socioeconòmics.
 - Qualitat ambiental.
- Creació i gestió d'una base de dades provinent de l'anomenat *Real World Data*. És a dir, informació secundària provinent de l'assistència primària i especialitzada (incloses hospitalitzacions i urgències) a la comarca de l'Alt Empordà.
- Desenvolupar mètodes estadístics per a controlar la qualitat d'aquesta base de dades.
- Integrar aquesta base de dades del *Real World Data* amb l'observatori de salut.
- Explotar la base integrada d'indicadors generant informes descriptius (taules i gràfics, taules creuades, prevalença i incidència, taxes estandarditzades), mapes (representació gràfica de les taxes estandarditzades, factors socioeconòmics, qualitat ambiental) i altres documents tècnics.

3 MÈTODES

3.1 DISSENY D'UN REPOSITORI D'INDICADORS DE SALUT, SOCIOECONÒMICS I MEDIAMBIENTALS

En aquesta secció es descriu quin és el procediment que s'ha dut a terme per a l'elaboració del repositori d'informació amb indicadors de salut, socioeconòmics i mediambientals. La secció està dividida en tres subseccions. En la primera es descriu el procediment que s'ha utilitzat per tal de definir els requeriments del repositori i en la segona, el disseny que s'ha fet servir per a construir-lo i finalment s'explica la manera en què s'ha construït.

3.1.1 Identificació de requeriments

Un repositori és un sistema informàtic on s'emmagatzema la informació d'una organització amb la finalitat que els membres d'una organització (sigui de persones o ens) puguin compartir-la de manera organitzada i estandarditzada. El repositori Indika de dades proporciona accés a indicadors d'interès sobre el territori de l'Alt Empordà.

La identificació de requeriments és el procés en el qual s'identifiquen les relacions funcionals, enteses com el conjunt de característiques que componen el producte global, amb la intenció d'identificar quins són els components que conformen el tot. En definitiva, un procés d'identificació de requeriments és una elaboració desglossada de les estructures que conformaran el sistema global amb la intenció de definir quins són els objectius de cadascun dels components. La intenció és dividir un sistema de tal forma que cada component es pugui descriure sense necessitat de referir a un altre component. D'aquesta manera, cada part del sistema tindrà funcions independents que poden reutilitzar-se i reemplaçar-se.

L'objectiu del repositori és facilitar el procés de reutilització de dades obertes rellevants sobre el territori per part dels diferents agents que el componen. La guia estratègica per a la posada en marxa de conjunts de dades obertes mínimes a publicar (Federación española de municipios y provincias, Red de entidades

Mètodes

locales por la transparencia y participación ciudadana, Wolter Kluwer, 2018) defineix les dades obertes com a dades amb una llicència que permet la seva reutilització.

El Pol de Salut i Social de l'Alt Empordà té, entre altres objectius, crear una xarxa entre entitats i organitzacions que treballin de forma conjunta per a la millora social i sanitària de la comarca. Per tant, un dels requeriments que s'ha tingut en compte és la necessitat de col·laboració entre diverses entitats entre les quals s'inclou l'observatori social de l'Alt Empordà i la mateixa Fundació Salut Empordà.

A patir de la discussió duta a terme amb els integrants d'Indika i la revisió bibliogràfica de múltiples observatoris i repositoris s'han decidit els següents requeriments generals: (1) les dades han de ser obertes, (2) donats els recursos inicials d'Indika cal reduir la necessitat de recursos i per tant, facilitar la col·laboració entre entitats, (3) ha de facilitar la producció científica, (4) les dades han d'estar auto contingudes, el que significa que es puguin fer servir de forma individual, (5) les dades s'han de poder agrupar en temàtiques concretes d'interès i (6) les dades han de ser d'accés obert.

Les dades obertes són aquelles dades que són accessibles i reutilitzables sense permisos específics. L'entitat *Open Knowledge Foundation* defineix les dades obertes com dades que poden ser utilitzades, reutilitzades i redistribuïdes lliurement per qualsevol persona, i que es troben subjectes, al requeriment d'atribuir i compartir de la mateixa manera en què apareixen, respectant la seguretat i privacitat de la informació sempre.

El procés per a identificar els requeriments s'ha basat en la revisió del conjunt d'indicadors presentats en l'informe *Key Figures on Europe – Statistics Illustrated* publicat de forma anual per EUROSTAT, el *Marco de indicadores mundiales para los Objetivos de Desarrollo Sostenible y metas de la Agenda 2030 para el Desarrollo Sostenible* (ONU, 2020) i l'*European Core Health Indicators shortlist of 88 health indicators identified by policy area* (European Commission).

Mètodes

3.1.1.1 Revisió d'indicadors

La revisió d'indicadors ha consistit en la cerca d'indicadors disponibles per a la comarca de l'Alt Empordà i, sempre que ha estat possible, també per als seus municipis. Per dur a terme la cerca s'han fet servir els indicadors descrits a *Key figures on Europe — Statistics illustrated — 2019 edition* (EUROSTAT, 2019).

S'ha creat una taula amb el codi de l'indicador, el nom de l'indicador i la descripció. Per a cada un dels indicadors s'ha indicat la seva disponibilitat a nivell Europeu, Espanyol, Català, Provincial, Comarcal i Municipal. Segons el nivell de desagregació desitjat s'ha fet una consulta a la web d'EUROSTAT, INE o IDESCAT en primera instància.

Un cop identificats els indicadors no disponibles de forma pública s'ha fet una consulta a Idescat. Feta aquesta consulta s'ha procedit a demanar la informació a les entitats responsables (segons la resposta d'Idescat).

3.1.1.2 Requisits tècnics

En aquest apartat es descriuen els requisits tècnics que ha de complir el repositori.

El repositori ha de ser un magatzem d'informació impulsat per la comunitat. Tenint en compte la filosofia d'Indika cal que aquest s'adapti a les necessitats dels diferents agents del territori. Per tant aquests agents han de ser capaços de decidir com volen que sigui aquest repositori i quina informació ha de contenir, així com participar de la recollida i actualització de les dades.

Donada la naturalesa de l'observatori és imprescindible que el seu repositori sigui descentralitzat. Això significa que no hi ha un únic mantenidor sinó que són totes les persones i entitats les responsables de mantenir-lo actiu i actualitzat. D'aquesta manera si una entitat vol deixar de participar en el projecte, la resta pot continuar a partir de la feina realitzada. De la mateixa manera, si una empresa o entitat vol fer-ne ús de forma local ha de tenir la capacitat de descarregar el repositori i fer-ne ús sense necessitat d'estar connectada a la xarxa.

Mètodes

El repositori ha de tenir la capacitat d'incloure nous indicadors en el futur de manera fàcil. Ha de ser fàcilment extensible. Per tal de complir aquest objectiu el repositori estarà format per petits fitxers on cada fitxer contindrà informació relacionada amb un indicador de manera que incloure un nou indicador és una tasca trivial.

Volem que les entitats que hi participen rebin el crèdit adequat. Per aquesta raó cal que el repositori registri qui envia cada indicador, així com qui és el responsable de la seva actualització. És a dir, cal assegurar la traçabilitat de la informació.

El repositori ha de ser obert. Les dades que en ell s'inclouen han de ser accessibles per tothom. D'aquesta manera cal permetre la correcció de la informació per garantir la fiabilitat d'aquest.

Donada la capacitat de fer un seguiment de les actualitzacions (traçabilitat) és més fàcil que cada persona rebi el crèdit que es mereix.

3.1.2 Disseny i arquitectura del repositori

Un cop identificats els requisits mínims del repositori, s'ha procedit a dissenyar la seva arquitectura tenint en compte les necessitats de l'entitat Indika i les seves característiques tant internes com externes.

3.2 CONSTRUCCIÓ D'UN REPOSITORI D'INFORMACIÓ SANITÀRIA, SOCIOECONÒMICA I MEDIAMBIENTAL DE L'ALT EMPORDÀ

Un cop identificats els requeriments de l'observatori s'ha optat per fer la seva implementació com un repositori de GitHub (GitHub, 2020). GitHub és una plataforma de desenvolupament col·laboratiu que permet allotjar projectes de codi obert i tancat amb un sistema de control de versions. Un sistema de control de versions és un sistema que s'encarrega de registrar els canvis que ha sofert un arxiu al llarg del temps amb la intenció de poder recuperar una versió específica en un moment temporal posterior. GitHub fa ús del sistema de control de versions Git (Chacon i Straub, 2014).

Mètodes

El repositori és un projecte de GitHub amb llicència GPL-3 (Free Software Foundation, 2020). La llicència GPL-3 és un tipus de llicència per a programari que permet la còpia, distribució (comercial o no) i modificació del codi, sempre que qualsevol modificació es continuï distribuint amb la mateixa llicència GPL. La llicència GPL no permet la distribució de programes executables sense el codi font corresponent o una oferta de com obtenir-lo gratuïtament. És la llicència lliure de programari més utilitzada.

El repositori està dividit en dos directoris: *indicators* i *indicators-list*. El directori *indicators* conté les dades i descripcions dels indicadors. Cada indicador es troba en un fitxer en format JSON. El directori *indicators-list* conté llistes d'indicators. Les llistes d'indicators són fitxers en format JSON que agrupen un conjunt d'indicators segons la seva temàtica o interès. Per exemple, la llista d'indicators 01-indicadors-clau-de-el-alt-emporda.json conté el llistat d'indicators utilitzats en el document "Indicadors Clau de l'Alt Empordà" (Franquet *et al.*, 2020). Les llistes d'indicators tenen com a funció facilitar la descàrrega de conjunts d'indicators i compartir agrupacions d'indicators per temàtiques d'interès.

Els indicadors estan dividits en dos atributs: meta i data. L'atribut meta és un objecte amb sis atributs que descriuen la informació continguda a data. La Taula 5 mostra els atributs que ha de contenir la meta informació.

Taula 5 Meta informació dels indicadors al repositori

Atribut	Descripció	Tipus
name	El nom de l'indicador. Ha de ser un nom idèntic al del fitxer	caràcter
Title	El nom de l'indicador	caràcter
Description	La descripció de l'indicador	caràcter

Mètodes

Source	La font de la qual s'han extret les dades. Sempre que sigui possible l'enllaç directe	caràcter
maintainer	La persona encarregada de mantenir l'indicador al repositori	caràcter
lastUpdate	Data de l'última actualització de l'indicador al repositori.	Data i temps

Font: Elaboració pròpia

3.2.1 API

Una interfície de programació d'aplicacions (API per les seves sigles en anglès) és una descripció del protocol de comunicació entre dos agents (normalment una màquina i un humà o dues màquines). Una API té com a funció facilitar i estandarditzar la forma en què els agents es relacionen amb el programari en qüestió. Un exemple d'API són les llibreries de R que faciliten l'aplicació de metodologies estadístiques mitjançant funcions concretes.

3.3 EXPLOTACIÓ DE LA BASE DE DADES INTEGRADA

3.3.1 Indicadors Clau de l'Alt Empordà

El document "Indicadors Clau de l'Alt Empordà" és un recull d'indicadors de diverses temàtiques com la població, la economia i el medi ambient a l'Alt Empordà. Aquest serà un document amb freqüència anual que té com a objectiu facilitar una visió global de la situació a l'Alt Empordà. El document està estructurat seguint el document d'EUROSTAT "*Key Figures on Europe – Statistics Illustrated 2019 version*" (EUROSTAT, 2019). Per realitzar aquest document el primer pas ha estat fer la revisió dels indicadors proporcionats per

Mètodes

l'EUROSTAT i avaluar la seva disponibilitat en diferents nivells territorials com son Espanya, Catalunya, la província de Girona i l'Alt Empordà.

3.4 IMPLEMENTACIÓ D'UNA LLIBRERIA DE R PER AL MODELATGE DE DADES BASEDES EN ESDEVENIMENTS

Aquesta secció està dividida en dos subapartats. El primer descriu que és una arquitectura de dades, que és una arquitectura basada en esdeveniments i quines parts ha de tenir una arquitectura basada en esdeveniments. En la segona part es descriu la implementació realitzada en el paquet eventr.

Un arquitectura o model de dades és un model abstracte sobre quina informació descriu un conjunt de dades i com està organitzada aquesta informació de forma detallada (Simsion i Graham, 2004).

Una arquitectura dirigida per esdeveniments (EDA) fa ús dels esdeveniments per accionar un conjunt d'accions. En la forma més simple té tres elements clau: l'esdeveniment, el manipulador i el repartidor. L'esdeveniment és l'encàrrec de transportar la informació necessària per realitzar una acció. El manipulador processarà la informació de l'esdeveniment per executar alguna tasca. El repartidor avaluarà els esdeveniments i decidirà quin manipulador ha d'utilitzar en funció del tipus d'esdeveniments. La Figura 11 mostra el diagrama de flux.

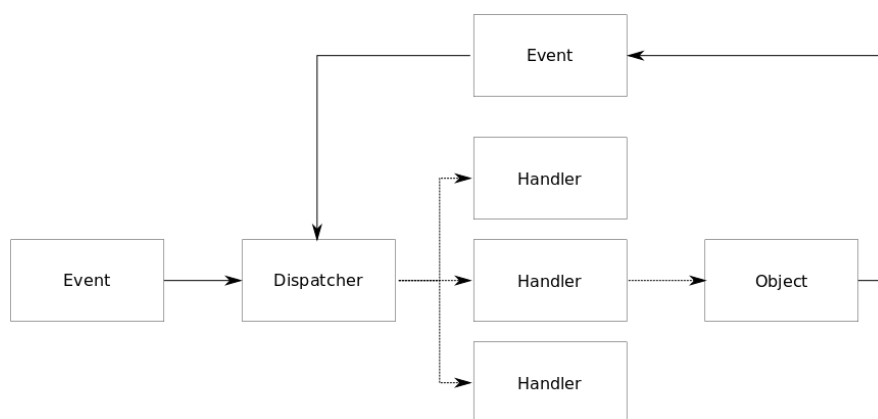


Figura 11: Diagrama de flux d'una arquitectura basada en esdeveniments. Font: Elaboració pròpia

3.4.1 Terminologia

A continuació es descriu la terminologia emprada:

Mètodes

- Els esdeveniments són objectes amb una capçalera i un cos. La capçalera emmagatzema tota la informació relativa al mateix esdeveniment com, per exemple, el seu tipus (`type`), una marca temporal (`time`) i un identificador (`id`). El cos emmagatzema la informació que descriu què ha passat durant aquest esdeveniment.
- L'especificació d'un esdeveniment és la definició de la seva estructura. L'especificació indica quina informació conté i com ha d'estar estructurada aquesta.
- L'ocurrència d'un esdeveniment és cada un dels successos que han tingut lloc d'una especificació.

A tall d'exemple suposem que ens interessa definir l'esdeveniment Naixement. L'especificació de l'esdeveniment Naixement defineix aquest esdeveniment com un objecte que té tres paràmetres: un identificador (`id`), la data de naixement (`birth_date`) i el lloc de naixement (`birth_place`). L'ocurrència de l'esdeveniment Naixement seria cada un dels naixements del nostre registre. La Taula 6 mostra exemples d'ocurrències de l'esdeveniment Naixement.

Taula 6 Exemple d'ocurrències de tres esdeveniments. Font: Elaboració pròpia.

id	birth_date	birth_place
eduardo-punset	1936-11-09	Barcelona
serj-tankian	1967-08-21	Beirut
ara-malikian	1968-09-14	Beirut

- Els Manipuladors són els encarregats d'aplicar els processos necessaris a un esdeveniment per transformar els seus paràmetres en els atributs d'un objecte.
- Els Repartidors són els encarregats d'enviar esdeveniments als seus respectius manipuladors.

Mètodes

3.4.2 Implementació

A continuació, es descriu la implementació realitzada en el paquet `eventr`. En aquesta implementació s'han inclòs tres classes principals: `event`, `handler` i `dispatcher`. Seguint el model proposat per Wickham (2019) cada classe té un constructor, un validador i un ajudant.

La classe `event` és una llista formada per dos elements: una capçalera (`header`) i un cos (`body`). La capçalera emmagatzema la informació relativa a l'esdeveniment. Concretament un identificador (`id`) de l'esdeveniment, temps (`time`) en què ha tingut lloc i el seu tipus (`type`). L'element `body` emmagatzema la informació que servirà per transformar els atributs d'un objecte d'interès.

Per facilitar la manipulació i l'ús dels esdeveniments s'han implementat les següents funcions.

- `get_body()`: Retorna la llista d'elements emmagatzemats en el cos de l'esdeveniment.
- `get_body_attr()`: A diferència de la funció `get_body()`, aquesta únicament retorna aquell atribut que coincideix amb el paràmetre `attr`.
- `get_header()`: Retorna una llista amb informació de la capçalera de l'esdeveniment.
- `get_id()` , `get_time()` i `get_type()`: Retornen l'identificador, el temps de l'esdeveniment i el tipus d'esdeveniment, respectivament.

Els esdeveniments poden agrupar-se en una llista d'esdeveniments `event_list`. Les llistes d'esdeveniments es poden crear utilitzant la funció `event_list()` o mitjançant l'operador `+`. D'aquesta manera, és possible definir una llista d'esdeveniments com la suma de tots els esdeveniments.

```
R> my_events_list <- first_event + second_event + nth_event
```

La funció `handler()` retorna un objecte de tipus `handler`. Els `handler` són els encarregats de modificar l'estat d'un objecte a partir de la informació continguda en un esdeveniment. La funció `handler()` té dos paràmetres: `type`

Mètodes

i `FUN`. El primer serveix per a indicar quin tipus d'esdeveniments ha de manipular el `handler`. `FUN` és la funció encarregada de modificar l'estat de l'objecte a partir de la informació continguda en l'esdeveniment. La funció `FUN` ha de ser del tipus:

```
R> foo = function(obj, event) {  
  obj <- do_something(event)  
  return(obj)  
}
```

On `obj` és l'objecte a modificar i `event` l'esdeveniment entrant. És a dir, la funció ha de tenir com a paràmetres d'entrada un objecte (`obj`) i un esdeveniment (`event`) i retornar únicament un objecte (`obj`) modificat.

S'han implementat les següents funcions per a la classe `handler`:

- `get_type()`: Indica per quin tipus d'esdeveniment és aquest `handler`.
- `get_fun()`: Retorna la funció (`FUN`) d'un objecte tipus `handler`.
- A l'igual que els esdeveniments, els manipuladors també es poden agrupar en llistes. La funció `handlers_list()` és l'encarregada de crear-les. Les `handlers_list` únicament poden contenir un manipulador per tipus d'esdeveniment. Si introduïm un manipulador per a un tipus d'esdeveniment que ja té un assignat, el segon sobreescriu al primer.

```
first_handler <- handler(  
  type = 'FIRST_EVENT_TYPE',  
  FUN = function(obj, event) return(obj)  
)  
  
second_handler <- handler(  
  type = 'SECOND_EVENT_TYPE',  
  FUN = function(obj, event) return(obj)  
)
```

Mètodes

```
type = 'SECOND_EVENT_TYPE',  
  
FUN = function(obj, event) return(obj)  
  
)  
  
my_handlers_list <- handlers_list(first_handler, second_handler)
```

També és possible fer servir l'operador + per crear una llista de manipuladors:

```
my_handlers_list <- first_handler + second_handler
```

La classe `dispatcher` és una llista formada per una llista de `handlers` i la funció `dispatch()`. Donada una llista d'esdeveniments, la funció `dispatch()` reparteix de forma seqüencial aquests esdeveniments als seus corresponents `handlers`. La funció `dispatch()` té tres paràmetres d'entrada: `obj`, `events` i `accumulate`. `obj` és l'objecte d'interès que serà modificat pels esdeveniments. `events` és una llista d'esdeveniments i `accumulate` és un booleà que indica si l'output de la funció ha de ser l'estat final de l'objecte o l'estat després de cada un dels esdeveniments inclosos en la llista d'esdeveniments (`events`).

Per a la classe `dispatcher` s'han implementat els següents mètodes:

- `add_handler()` o `+`: Permeten afegir un manipulador a un repartidor.
- `get_handlers()`: Permet recuperar la llista de manipuladors inclosos en un repartidor.

`get_dispatch()`: Retorna la funció `dispatch()` creada a partir dels manipuladors.

4 RESULTATS

En aquest capítol es presenten els resultats obtinguts a conseqüència de l'activitat duta a terme en el marc d'aquesta tesi. El capítol està dividit en dues parts. La primera part correspon a l'anàlisi de l'impacte que ha tingut l'observatori en el territori. La segona part correspon a l'anàlisi de l'impacte que ha tingut la llibreria eventr.

4.1 INDIKA POL DE SALUT I SOCIAL DE L'ALT EMPORDÀ

Un dels objectius principals del projecte és avaluar la capacitat d'explotar la informació integrada per millorar l'estat de benestar de la comarca. Un dels indicadors emprats per a avaluar aquesta capacitat és el nombre de projectes elaborats. En total s'han realitzat 5 infografies, 1 aplicació web sobre la situació de la COVID-19 a l'Alt Empordà, l'informe "Indicadors Clau de l'Alt Empordà" (Franquet *et al.*, 2020) i la pàgina web d'Indika (Indika Pol de Salut i Social de l'Alt Empordà, 2020).

4.1.1 Infografies

En aquesta secció es descriuen les infografies elaborades des de l'observatori. S'han realitzat tres infografies: Població a l'Alt Empordà, La natalitat a l'Alt Empordà i Un any de COVID-19 a l'Alt Empordà.

4.1.1.1 Població a l'Alt Empordà

La primera infografia creada correspon a la infografia "Població de l'Alt Empordà" (Fundació Salut Empordà, 2019). L'estratègia adoptada té com a objectiu respondre a les preguntes: Quina és la població de l'Alt Empordà? Quina és la població pediàtrica? Quina és la població amb més de seixanta-quatre anys? Quina és la població sobre envellida? Entenent població sobre envellida com aquella amb més de vuitanta-quatre anys. La infografia pretén emmarcar les dades en el seu context geogràfic fent la comparació amb la informació disponible per al conjunt de Catalunya.

Les dades utilitzades per a la infografia provenen dels estadístics: “Estructura per edats, envelliment i dependència” (Idescat, 2020a) i “Població a 1 de gener. Per sexe i edat any a any” (Idescat, 2021a).

La transformació i explotació de les dades s’ha realitzat mitjançant el llenguatge de programació R (R Core Team, 2020). El codi corresponent es pot trobar a l’annex *Codi de la infografia “Població a l’Alt Empordà”* d’aquesta tesi.

Un cop transformades les dades i elaborada una anàlisi descriptiva de les mateixes es realitzen diversos esbossos. La Figura 12 mostra tres dels esbossos realitzats.

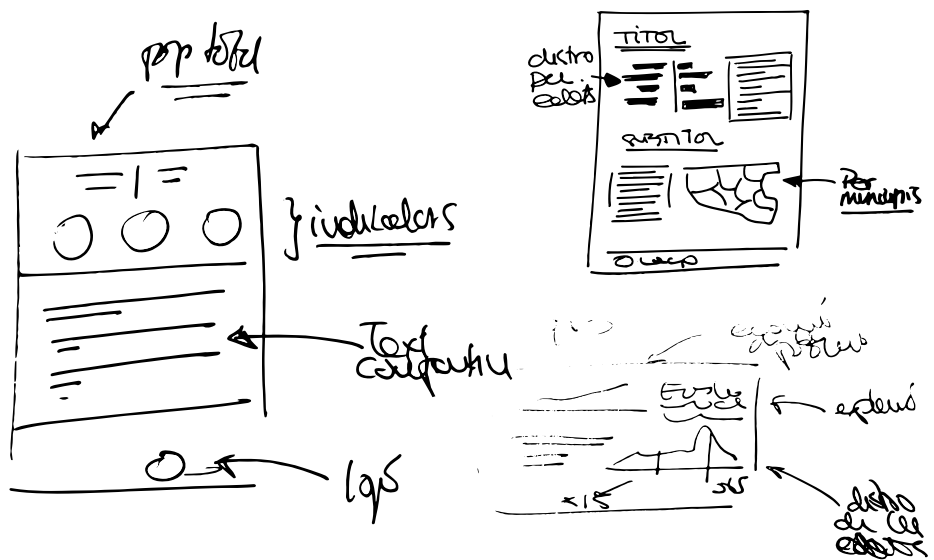


Figura 12 Tres esbossos per a la infografia “Població a l’Alt Empordà”. Elaboració pròpia.

Un cop s’ha realitzat l’esbós i se selecciona quina de totes les opcions és la més adequada es realitza un prototip amb les dades reals i finalment es revisa el text i les diferents explicacions. La Figura 13 mostra el resultat final que també es pot trobar a la pàgina web d’Indika (Fundació Salut Empordà, 2019) o bé a l’annex Infografia “Població a l’Alt Empordà”.



Figura 13 Infografia "Població a l'Alt Empordà" versió 2019. Elaboració pròpia.

4.1.1.2 infografia "La Natalitat a l'Alt Empordà"

La infografia "La Natalitat a l'Alt Empordà" té com a objectiu mostrar la informació relativa als naixements produïts l'any 2020. La natalitat d'una regió és un dels valors clau en el creixement demogràfic i per aquesta raó és important analitzar-la.

La visualització té com a objectiu respondre a les preguntes: Que ha passat amb la natalitat a l'Alt Empordà? És a dir, quina ha estat l'evolució de la natalitat a l'Alt Empordà. Quina és l'estructura de la natalitat en l'actualitat i per quina raó té aquesta estructura? En relació amb els ODS la natalitat que tenim a l'Alt Empordà és una natalitat sostenible i saludable?

Les dades emprades per a la realització de la visualització s'han obtingut de registres propis de la Fundació Salut Empordà, de la taxa bruta de natalitat (Idescat, 2021b), naixements per edats de la mare (Idescat, 2021c). El codi amb el qual s'han explorat i transformat les dades es pot trobar a l'annex Codi de la infografia "Natalitat a l'Alt Empordà".

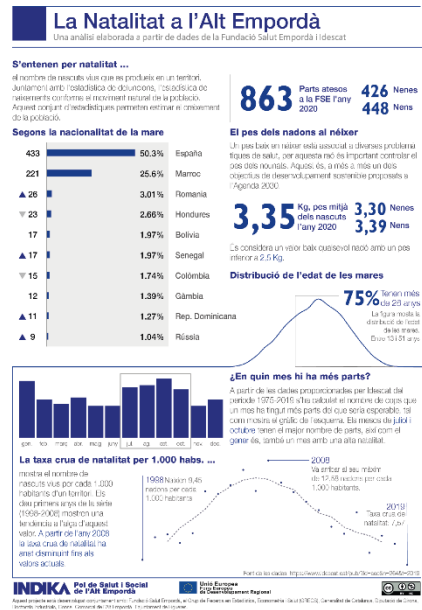
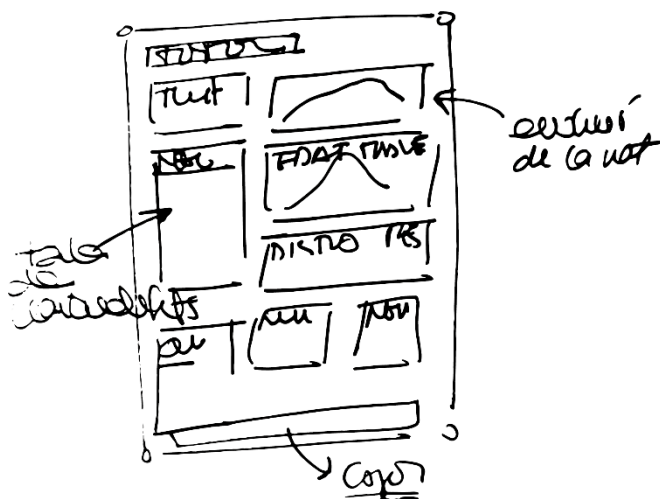


Figura 14 Esbós de la infografia “Natalitat a l’Alt Empordà” i infografia final. Elaboració pròpia.

La Figura 14 mostra un dels esbossos realitzats i la infografia final la “Natalitat a l’Alt Empordà” que també es pot trobar a l’annex Infografia “Natalitat a l’Alt Empordà” o bé a la pàgina web d’Indika (Indika Pol de Salut i Social de l’Alt Empordà, 2021).

4.1.1.3 Infografia “Un any de COVID-19 a l’Alt Empordà”

Degut a la situació provocada per la pandèmia de la COVID-19 i donat l’objectiu de l’observatori s’han executat diverses accions amb la intenció d’informar a la població sobre l’estat de la comarca en aquest aspecte. Concretament s’ha fet a través d’una aplicació web que es descriu a l’apartat 4.1.2 i mitjançant la infografia “Un any de COVID-19 a l’Alt Empordà” on s’analitza amb més profunditat quin ha estat el comportament de la pandèmia al llarg de l’any.

L’objectiu de la infografia és mostrar quina ha sigut l’evolució de la pandèmia a la comarca i si hi ha hagut diferències entre el comportament que ha tingut aquesta a l’Alt Empordà i a Catalunya així com les diferències en l’estat de la vacunació.

Per realitzar la infografia s’han emprat dades de diverses fonts d’informació. S’ha utilitzat el “Registre de casos de COVID-19 realitzats a Catalunya. Segregació per sexe i municipi” (Departament de Salut de la Generalitat de Catalunya,

2021a), Registre de casos de COVID-19 realitzats a Catalunya. Segregació per sexe i edat” (Departament de Salut de la Generalitat de Catalunya, 2021b), el “Registre de defuncions per COVID-19 a Catalunya. Segregació per sexe i comarca” (Departament de Salut de la Generalitat de Catalunya, 2021c), “Vacunació per al COVID-19 per comarca” (Departament de Salut de la Generalitat de Catalunya, 2021d) i el nombre d’ingressos hospitalaris per COVID-19 a la FSE.

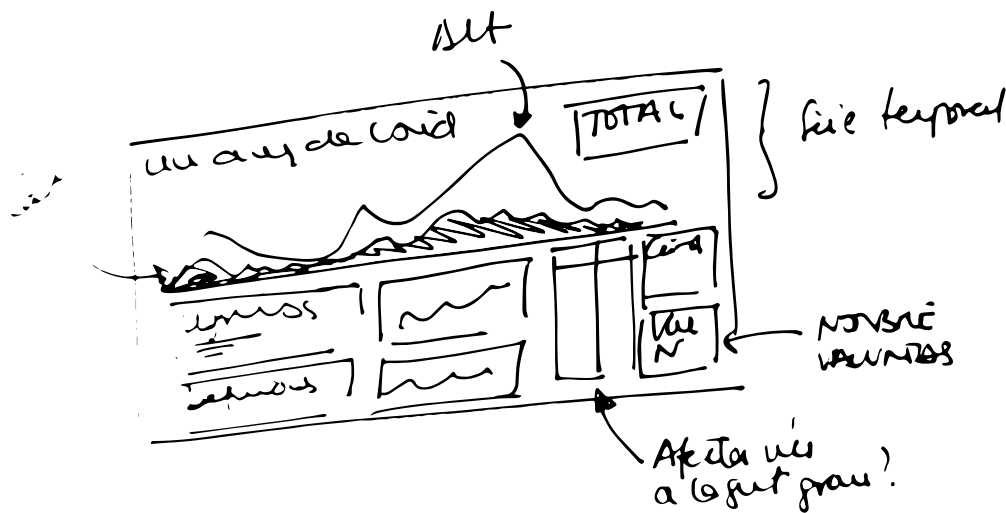


Figura 15 Esbós de la infografia "Un any de COVID-19 a l'Alt Empordà". Font: Elaboració pròpia.

La Figura 15 mostra l’esbós per al disseny de la infografia i la Figura 16 mostra un dels primers prototips a partir del qual es decideix modificar l’estructura de la infografia per tal de fer-ho en dues pàgines.

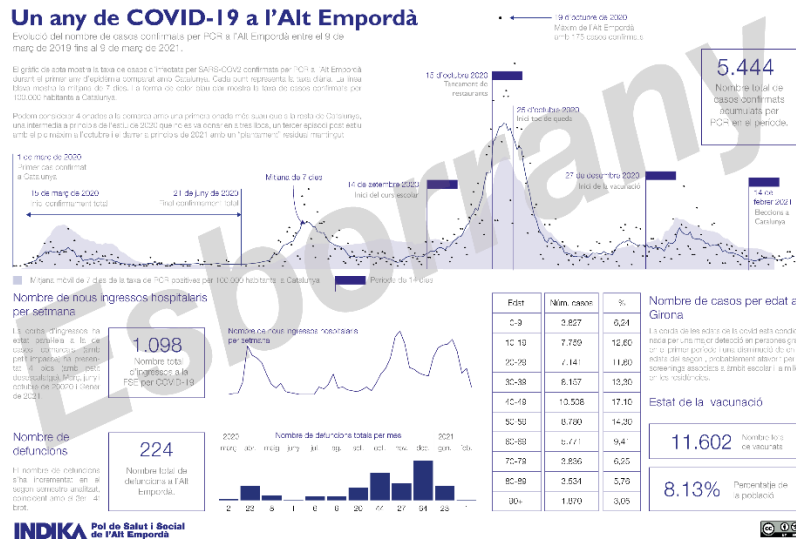


Figura 16 Prototip de la infografia "Un any de COVID-19 a l'Alt Empordà". Font: Elaboració pròpia.

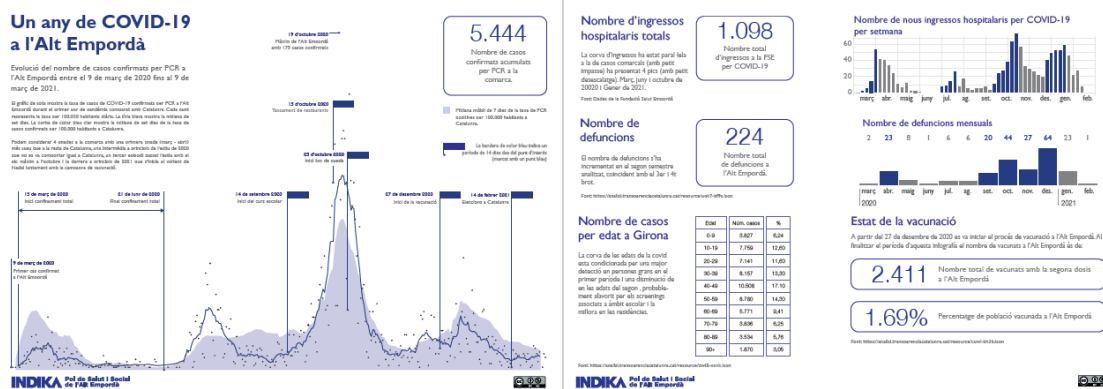


Figura 17 Infografia "Un any de COVID-19 a l'Alt Empordà". Font: Elaboració pròpia.

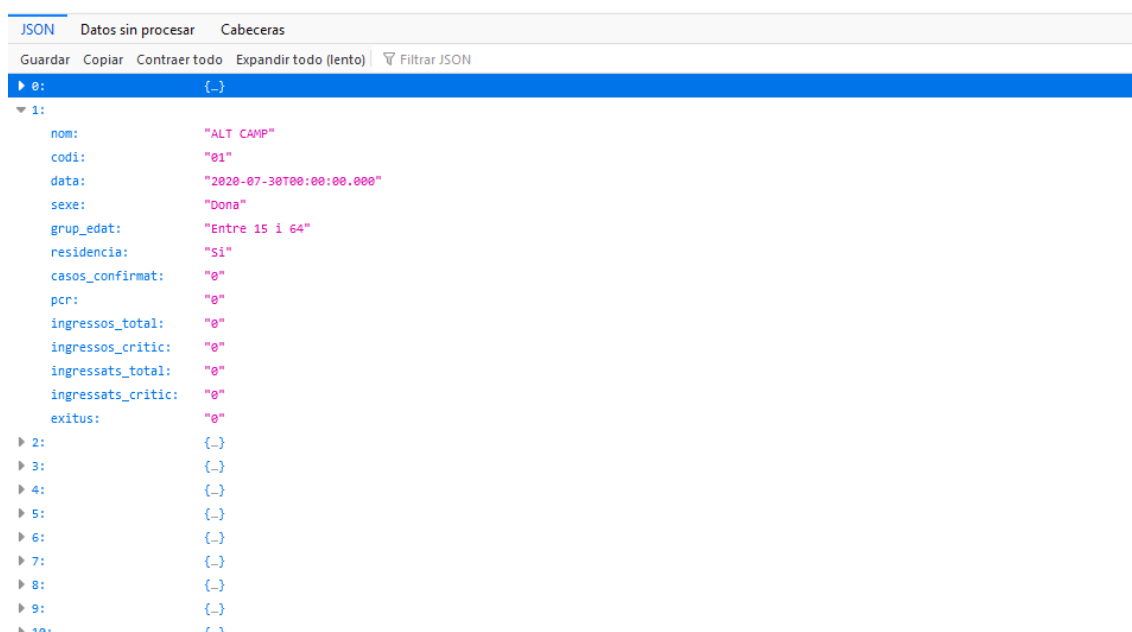
La Figura 17 mostra la infografia final també disponible a l'Annex Infografia "Un any de COVID-19 a l'Alt Empordà" juntament amb el codi R emprat per a la seva elaboració.

4.1.2 Aplicació sobre la situació de la COVID-19 a l'Alt Empordà

Donada la situació de pandèmia mundial provocada per la COVID-19 i, tenint en compte els objectius d'un observatori de salut, des de l'observatori Indika s'ha optat per crear una aplicació web amb la informació disponible i actualitzada sobre l'estat de la comarca de l'Alt Empordà en el referent a la pandèmia. Aquesta aplicació té com a objectiu proporcionar informació rellevant a la població i als agents territorials per millorar la presa de decisions i emprendre les accions necessàries. L'aplicació obté la informació mitjançant l'API

proporcionada pel portal de dades obertes Catalunya (Generalitat de Catalunya, 2020).

L'API per a l'obtenció de les dades és un servei web que permet als desenvolupadors fer consultes a la base de dades mitjançant el protocol de transferència d'arxius cURL (Stenberg, 2020). Aquest protocol fa ús d'un localitzador uniforme de recursos (URL per les seves sigles en anglès) per indicar a quin conjunt de dades es vol accedir. Per exemple, si accedim, mitjançant un navegador, a la URL <https://analisi.transparenciacatalunya.cat/resource/c7sd-zy9j.json> accedirem a una mostra de les dades diàries de COVID-19 per comarca (Departament de Salut, 2020) en format JSON. La Figura 18 mostra el resultat d'accedir a l'enllaç mencionat.



```
JSON  Datos sin procesar  Cabeceras
Guardar Copiar Contraer todo Expandir todo (lento) Filtrar JSON
▶ 0: {}
▼ 1:
  nom: "ALT CAMP"
  codi: "01"
  data: "2020-07-30T00:00:00.000"
  sexe: "Dona"
  grup_edat: "Entre 15 i 64"
  residencia: "Si"
  casos_confirmat: "0"
  pcr: "0"
  ingressos_total: "0"
  ingressos_critic: "0"
  ingressats_total: "0"
  ingressats_critic: "0"
  exitus: "0"
▶ 2: {}
▶ 3: {}
▶ 4: {}
▶ 5: {}
▶ 6: {}
▶ 7: {}
▶ 8: {}
▶ 9: {}
▶ 10: {}
```

Figura 18 Captura de pantalla del fitxer de dades diàries de Covid-19 per comarca del Departament de Salut. Elaboració pròpia.

Aquest protocol de transferència d'arxius permet fer consultes específiques afegint paràmetres a l'URL. En el conjunt de dades diàries de COVID-19 per comarca (Departament de Salut, 2020) podem especificar quina comarca ens interessa fent ús del paràmetre codi. Així, si volem accedir a les dades de l'Alt Empordà faríem servir l'URL:

<https://analisi.transparenciacatalunya.cat/resource/c7sd-zy9j.json?codi=02>

on codi = 02 és l'identificador de la comarca "Alt Empordà". Cada conjunt de dades té uns paràmetres específics. Podem consultar les opcions disponibles a la seva documentació (Generalitat de Catalunya, 2020). La Taula 7 Dades diàries de COVID-19 per comarca" mostra els camps que componen la taula amb informació diària sobre COVID-19 a les comarques de Catalunya a partir de la qual s'han fet els càlculs d'ingressos i èxits. La informació d'aquest conjunt de dades s'actualitza amb tres dies de retard aproximadament.

Taula 7 Dades diàries de COVID-19 per comarca

Nom de la variable	Descripció	Tipus
NOM	Nom de comarca	Text Pla
CODI	Codi de comarca	Text Pla
DATA	Data	Data i temps
SEXE	Camp binari de desagregació de dades: inclou opcions HOME i DONA	Text Pla
GRUP_EDAT	Camp de desagregació de dades: inclou opcions: Menors de 15 anys Entre 15 i 64 anys Entre 65 i 74 anys Majors de 75 anys	Text Pla
RESIDENCIA	Camp binari de desagregació de dades: inclou opcions Si (població ingressada en residència geriàtrica) No (població no ingressada en residència geriàtrica)	Text Pla
CASOS_CONFIRMAT	Camp numèric amb dades de casos confirmats de Covid-19 (La PCR+ ens	Nombre

	indica que és un cas actiu i pel seguiment epidemiològic és important la data en què s'inicia el cas. No usem els casos COVID-19 confirmats per testos ràpids serològics o per ELISA)	
PCR	Camp numèric amb dades de Proves PCR realitzades (La prova de la PCR per la detecció del SARS-CoV-2 es pot realitzar a diferents nivells: a l'atenció primària, hospitalària, laboral, salut pública... En la seva gran majoria, estan enregistrades a l'RSA. Des de l'ECAP es recuperen els resultats de les PCR que es fan en l'àmbit privat dels pacients que es visiten a l'atenció primària. Les PCR s'usen en general per confirmar un cas sospitós de COVID-19 o per fer el cribratge dels contactes d'un cas confirmat i per fer-ne el seguiment. El resultat de la prova pot trigar entre unes hores i uns dies en arribar, segons el lloc on es realitza. En la web el resultat de la PCR s'imputa a la data de realització de la prova, i la xifra pot anar augmentant retrospectivament a mesura que es van rebent els resultats dels laboratoris)	Nombre
INGRESSOS_TOTAL	Camp numèric amb dades d'ingressos registrats (Nombre de nous ingressos hospitalaris per COVID-19 que hi ha hagut durant el període descrit. Aquests	Nombre

	s'imputen al territori on està assignat el pacient a l'RCA)	
INGRESSOS_CRITIC	Camp numèric amb dades d'ingressos a UCI registrats (Nombre de nous ingressos a la UCI de pacients amb COVID-19 durant el període descrit. Aquests pacients estan inclosos en els ingressos, però en els dies previs (per ingressar a la UCI s'ha d'haver ingressat prèviament a l'hospital). Aquests s'imputen al territori on està assignat el pacient a l'RCA)	Nombre
EXITUS	Camp numèric amb dades de defuncions registrades (Les defuncions poden portar cert retard en la notificació. Les defuncions s'imputen al dia de la defunció i no el dia de la notificació. És per això que poden anar augmentant de forma retrospectiva)	Nombre

Font: Departament de Salut de la Generalitat de Catalunya (2021b)

També s'ha extret informació a partir del conjunt de dades "Registre de casos de COVID-19 realitzats a Catalunya. Segregació per sexe i municipi" (Departament de Salut de la Generalitat de Catalunya, 2021d) que conté informació actualitzada dia a dia sobre el nombre de casos a cada municipi. Per a la creació de l'aplicació només s'han fet servir aquells casos amb PCR positiva i no s'han tingut en compte els tests ràpids, ELISA o d'altres.

Taula 8 Estructura de la taula "Registre de casos de COVID-19 realitzats a Catalunya"

Nom de la variable	Descripció	Tipus
--------------------	------------	-------

TipusCasData	Data / Fecha / Date	Data i temps
ComarcaCodi	Codi de Comarca / Código de comarca / Region code	Text Pla
ComarcaDescripcio	Comarca / Comarca / Region	Text Pla
MunicipiCodi	Codi de municipi / Código de municipio / Municipality code	Text Pla
MunicipiDescripcio	Municipi / Municipio / Municipality	Text Pla
SexeCodi	Codi de sexe (0-home, 1-dona) / Código de sexo (0-hombre, 1- mujer) / Gender code (0- male, 1-female)	Text Pla
SexeDescripcio	Sexe / Sexo / Gender	Text Pla
TipusCasDescripcio	Resultat / Resultado / Result	Text Pla

NumCasos	Nombre de casos / Número de casos / Number of cases	Nombre
----------	---	--------

Font: Departament de Salut de la Generalitat de Catalunya (2021d)

Fent ús del llenguatge de programació R (R Core Team, 2020), la llibreria RSocrata (Devlin *et al.*, 2019) i la llibreria tidyverse (Wickham *et al.*, 2019) l'aplicació, "Situació de la COVID-19 a l'Alt Empordà" descarrega la informació d'interès i la transforma a un format adequat per al seu tractament posterior com fer els gràfics o mostrar el nombre de casos confirmats.

La interfície d'usuari (UI) i la lògica del servidor s'han implementat amb les llibreries Shiny (Chang *et al.*, 2020) de R i htmltools (Cheng *et al.*, 2020). Shiny és un *framework* per a la creació d'aplicacions web que fa ús d'un paradigma de programació reactiu. La interfície s'ha creat fent ús del marc de treball Bootstrap. Bootstrap és un *framework* que permet dotar a la interfície d'usuari d'una estètica concreta mitjançant llenguatge HTML de forma simple.

La interfície d'usuari (Figura 19 Aplicació "Situació de la COVID-19 a l'Alt Empordà") mostra un mapa de l'Alt Empordà, dividit per municipis, en la part central. A la dreta hi ha un panell que permet seleccionar quines dades s'han de mostrar al mapa. A la part superior del panell podem seleccionar si volem veure els nombres absoluts o les taxes per 100.000 habitants. Just a sota, podem seleccionar el marc temporal de les dades. El panell inferior mostra els nombres per al conjunt de l'Alt Empordà, Girona i Catalunya. A més a més mostra dos gràfics amb l'històric dels valors seleccionats. El superior mostra les dades crues i l'inferior mostra les dades acumulades.

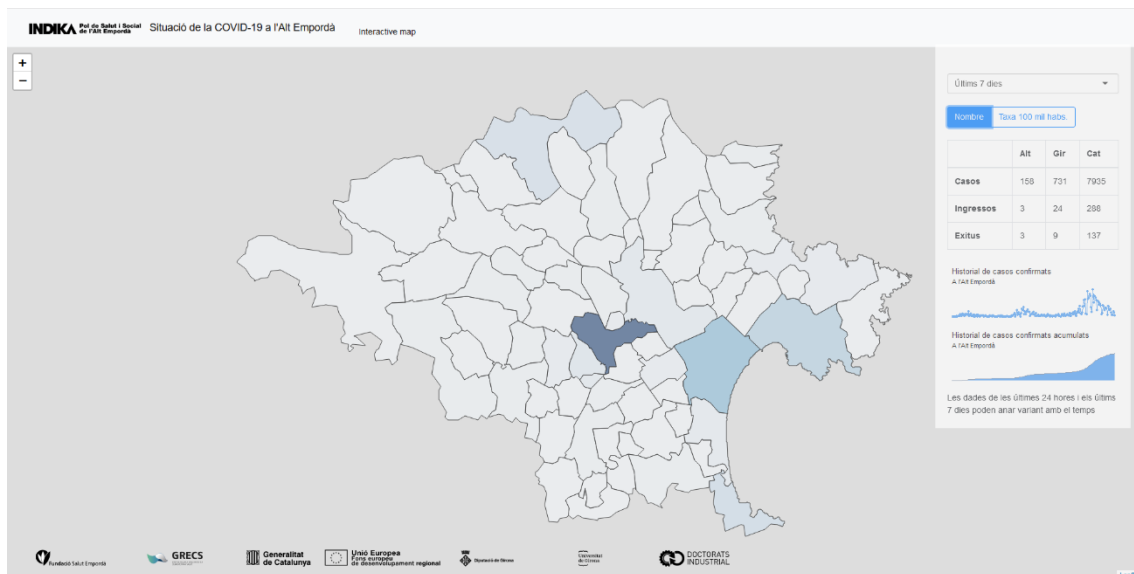


Figura 19 Aplicació "Situació de la COVID-19 a l'Alt Empordà". Font: Elaboració pròpia.

L'aplicació, que es pot trobar allotjada a la plataforma Shinyapps.io a l'enllaç: <https://fundacio-salut-emporda.shinyapps.io/covid-app/>, està estructurada seguint aquesta estructura:

- **Data:** El directori que emmagatzema la informació descarregada de la plataforma de dades obertes.
 - Altemporda.rds: Mapa dels municipis de l'Alt Empordà
 - Catalunya.rds: Conté els valors dels últims 14 dies sobre casos confirmats, ingressos hospitalaris i èxits a Catalunya per dies.
 - Comarca.rds: Conté les dades històriques per a la comarca de l'Alt Empordà sobre nombre de casos positius de COVID-19 confirmats per PCR positiva. .
 - Creds.rds: fitxer que emmagatzema les credencials de la plataforma de dades obertes per tal d'identificar qui en fa ús i controlar que aquest ús sigui adequat.
 - Girona.rds: Conté els valors dels últims 14 dies sobre casos confirmats, ingressos hospitalaris i èxits a la província de Girona per dies.
 - Municipis.rds: Conté informació dels últims 14 dies sobre el nombre de positius de COVID-19 confirmats per PCR positiva a l'Alt Empordà.

- Municipis.cat.rds: Conté el nombre de casos de COVID-19 confirmats al conjunt de Catalunya.
- Municipis-gir.rds: Conté el nombre de casos de COVID-19 confirmats al conjunt de la província de Girona.
-
- **R:**
 - 00-dades-per-comarca.R: Aquest fitxer és l'encarregat de descarregar la informació i guardar el fitxers al directori Data.
- **www:**
 - Css: Aquest directori conté fitxer per a millorar l'estètica de l'aplicació.
 - Img: aquest directori conté les imatges utilitzades a l'aplicació. En aquest cas conté els logotips de la fundació, el pol de salut i la universitat entre d'altres
 - Js: Conté funcions per fer que el mapa sigui interactiu i mostri informació quan es fa clic sobre algun dels municipis
- **App.R:** L'arxiu app.R conté el codi R encarregat de controlar el servidor i de mostrar la informació en pantalla.

Covid-app.Rproj: Un arxiu que defineix el directori com un projecte de R.

4.1.3 Indicadors Clau de l'Alt Empordà

Indicadors Clau de l'Alt Empordà (Franquet et al., 2020) és un recull de més de 30 indicadors de l'Alt Empordà que té com objectiu contextualitzar la informació del territori; ser la base per a la presa de decisions i a més ha de servir com a eina per fer conèixer l'observatori. El document complet es pot trobar a la pàgina web d'Indika.

El disseny d'aquest document s'ha fet tenint com a referència el document creat per Eurostat "*Key Figures on Europe*" (2019, 2020). Les dades han estat

recopilades de diferents fonts com la DGT, el departament d'Economia, Idescat i les bases de dades del mateix Hospital.



Figura 20 Portada d'Indicadors Clau de l'Alt Empordà 2020. Font: Elaboració pròpia

El document es pot trobar en català, castellà i anglès a la pàgina web d'Indika (<http://www.indikasalut.cat/>). En l'actualitat s'està treballant en la versió de l'any 2021.

4.1.4 Pàgina web del projecte Indika

La Taula 9 mostra les dades de relatives al trànsit de la pàgina web del projecte Indika entre els mesos del Gener fins al Juny. La columna "Nombre de visites" mostra el nombre de visites totals que ha tingut la pàgina web, la columna

“Nombre de visitants” indica el nombre de visitants diferents que han fet ús de la pàgina web, la columna “pàgines” indica el nombre de pàgines visitades.

Taula 9 Trànsit de la pàgina web d'Indika entre els mesos de Gener i Juny.

Any	Mes	Nombre de visites	Nombre de visitants	Pàgines
2021	January	108	203	2.405
2021	February	188	391	1.074
2021	March	311	679	4.732
2021	April	554	981	2.808
2021	May	563	1,110	2.608
2021	June	363	583	1.144

4.1.5 Altres mitjans

Com a resultat de l'activitat realitzada l'observatori ha aparegut diversos cops a la premsa i a mitjans de comunicació radiofònics. L'annex Recull de premsa presenta les diferents aparicions de l'observatori a la premsa.

4.2 LLIBRERIA EVENTR DE R

En aquesta secció es mostra com fer ús del paquet eventr per implementar un model de dades basat en esdeveniments que té com a objectiu estructurar informació mèdica d'un individu com la seva data de naixement, data de defunció i diverses dades antropomètriques.

El primer pas és definir quina informació cal emmagatzemar i com es vol emmagatzemar aquesta informació. Per a aquest exemple, es vol registrar la

següent informació: data de naixement, data de defunció i dades antropomètriques. Les dades antropomètriques consten de: alçada, pes i índex de massa corporal. L'índex de massa corporal es calcula a partir de l'alçada i el pes.

Un cop tenim clar quina informació necessitem cal definir l'estructura de l'objecte individu.

4.2.1 Estructura de l'individu

En aquest exemple es defineix l'objecte individual com una llista amb tres elements: `birth_date`, `death_date` i `anthropometric_measures`. L'element `anthropometric_measures` és una llista amb tres elements més: `weight`, `height` i `body_mass_index`.

```
R > library(eventr)
R > new_individual <- function() {
R +   return(
R +     list(
R +       birth_date = NA,
R +       death_date = NA,
R +       anthropometric_measures = list(
R +         weight = NA,
R +         height = NA,
R +         body_mass_index = NA
R +       )
R +     )
R +   )
R + }
R >
R > individual <- new_individual()
```

4.2.2 Crear les especificacions

Donats els objectius proposats cal crear, com a mínim, les especificacions de tres esdeveniments: l'esdeveniment Naixement, l'esdeveniment Defunció i l'esdeveniment Mesures antropomètriques. L'especificació d'un esdeveniment és una funció que retorna un objecte de tipus event.

```
R > specification <- function(<params>) {
R +   return(
R +     event(
R +       id = <event_id>,
R +       type = <event_type>,
R +       time = <event_time>,
R +       <event_params>
R +     )
R +   )
R + }
```

L'especificació de l'esdeveniment `add_birth_spec` té com a paràmetres `event_id` i `birth_date`. El paràmetre `event_id` és un caràcter que permet identificar un esdeveniment concret. Ha de ser un valor únic. En cas de no proporcionar un identificador, la funció `event()` assigna un identificador únic universal (`uuid` per si sigles en anglès) usant el paquet `uuid`. El paràmetre `birth_date` ha de ser una data en format `YYYY-MM-DD`. Per a aquest exemple el paràmetre `time` serà igual a la data de naixement `birth_date`.

```
R > add_birth_spec <- function(event_id, birth_date) {
R +   return(
R +     event(
R +       id = event_id,
R +       type = 'BIRTH',
R +       time = birth_date,
R +       birth_date = birth_date
R +     )
R +   )
R + }
```

L'especificació de l'esdeveniment `add_death_spec` té com a paràmetres `event_id` i `death_date`. El paràmetre `event_id` és igual al de l'especificació `add_birth_spec` i el paràmetre `death_date` és similar al paràmetre `birth_date` d'aquesta especificació.

```
R> add_death_spec <- function(event_id, death_date){
```

```

R+   return(
R+     event(
R+       id = event_id,
R+       type = 'DEATH',
R+       time = death_date,
R+       death_date = death_date
R+     )
R+   )
R+ }

```

L'especificació de l'esdeveniment `add_anthropometric_measures_spec` té tres paràmetres: `event_id` (igual als anteriors), `weight` i `height`. El paràmetre `weight` representa el pes de l'individu en quilograms. El paràmetre `height` representa l'alçada de l'individu en metres. Atès que l'índex de massa corporal es calcula a partir d'aquests dos valors no cal incloure-ho en l'especificació de l'esdeveniment. El `handler` serà l'encarregat de realitzar els càlculs.

```

R> add_anthropometric_measures_spec <-
R+ function(event_id, time, weight, height){
R+   return(
R+     event(
R+       id = event_id,
R+       time = time,
R+       type = 'ADD_ANTHROPOMETRIC_MEASURES',
R+       weight = weight,
R+       height = height R+   )
R+   )
R+ }

```

4.2.3 Crear els manipuladors

La funció dels manipuladors és transformar els atributs d'un objecte (en aquest cas l'objecte `individual`) a partir de la informació continguda en un esdeveniment. Per a cada tipus d'esdeveniment creat ha d'haver-hi un manipulador (`handler`).

El manipulador de l'esdeveniment Naixement (`birth_handler`) s'encarrega de modificar l'atribut `birth_date` de l'objecte individual i assignar-li el valor `birth_date` de l'esdeveniment entrant.

```
R> birth_handler <- handler(  
R+   type = 'BIRTH',  
R+   FUN = function(obj, event){  
R+  
R+     obj$birth_date <- get_body_attr(event, 'birth_date')  
R+     return(obj)  
R+  
R+   })
```

El manipulador de l'esdeveniment DEATH (`death_handler`) s'encarrega de modificar l'atribut `death_date` de l'objecte individual i assignar-li el valor `death_date` de l'esdeveniment entrant.

```
R> death_handler <- handler(  
R+   type = 'DEATH',  
R+   FUN = function(obj, event){  
R+  
R+     obj$death_date <-  
R+     get_body_attr(event, 'death_date')  
R+     return(obj)  
R+  
R+   })
```

El manipulador de l'esdeveniment `ADD_ANTHROPOMETRIC_MEASURES` (`anthropometric_measures_handler`) és una mica més complex. Aquest manipulador assigna els valors `weight` i `height` de l'esdeveniment als atributs `weight` i `height` de l'objecte individual i calcula, a partir d'ells, el valor de l'atribut `body_mass_index`.

```

R> anthropometric_measures_handler <- handler(
R+   type = 'ADD_ANTIROPOMETRIC_MEASURES',
R+   FUN = function(obj, event){
R+
R+     weight = get_body_attr(event, 'weight')
R+     height = get_body_attr(event, 'height')
R+     body_mass_index = weight / height^2
R+
R+     obj$anthropometric_measures$weight = weight
R+     obj$anthropometric_measures$height = height
R+     obj$anthropometric_measures$body_mass_index=
body_mass_index
R+
R+     return(obj)
R+   }
R+ )

```

4.2.4 Crear el repartidor

La funció `dispatcher()` crea, a partir dels manipuladores (`handlers`) indicats, un objecte de tipus `dispatcher`. Aquest objecte té dos elements: la llista de `handlers` i la funció `dispatch()` que és la funció encarregada d'enviar cada esdeveniment al seu respectiu manipulador.

```

R> repartidor <- dispatcher(
R+   handlers = handlers_list(
R+     birth_handler,
R+     death_handler,
R+     anthropometric_measures_handler
R+   )
R+ )
R>
R> str(repartidor, max.level = 1)

List of 2
 $ handlers:List of 3
  .. attr(*,"class")= chr [1:2] "handlers_list" "list"

```

```
$dispatch:function(obj = NULL, events, accumulate = FALSE)
- attr(*, "class")= chr [1:2] "dispatcher" "list"
```

4.2.5 Crear les ocurrencies dels esdeveniments

Les ocurrencies dels esdeveniments són aquells esdeveniments que han tingut lloc i han canviat l'estat de l'objecte individual. Per a aquest exemple s'han creat quatre ocurrencies. La primera correspon al naixement de l'individu, la segona i la tercera a dos moments en els quals s'han pres mesures antropomètriques i finalment, la defunció d'aquest.

```
R> the_events <-
R+   add_birth_spec(
R+   'first-event',
R+   birth_date = as.Date('1936-11-09')
R+ ) +
R+   add_anthropometric_measures_spec(
R+   'second-id',
R+   as.Date('1951-01-26'),
R+   65,
R+   1.72
R+ ) +
R+   add_anthropometric_measures_spec(
R+   'third-id',
R+   as.Date('1982-05-03'),
R+   70,
R+   1.72
R+ ) +
R+   add_death_spec(
R+   'fourth-id',
R+   death_date = as.Date('2018-05-11')
R+ )
```

4.2.6 Executar els esdeveniments

Un cop creats els esdeveniments podem calcular l'estat de l'objecte individual usant la funció `dispatch()` de l'objecte repartidor.

```
R> repartir <- get_dispatch(repartidor)
R> individual_01 <- repartir(obj = individual, events=the_events)
R> individual_01

$birth_date
[1] "1936-11-09"

$death_date
[1] "2018-05-11"

$anthropometric_measures

$anthropometric_measures$weight
[1] 70

$anthropometric_measures$height
[1] 1.72

$anthropometric_measures$body_mass_index
[1] 23.66144
```

Pot ser d'interès obtenir l'estat de l'individu en cada un dels esdeveniments. Per a això la funció `dispatch()` permet indicar usant el paràmetre `accumulate`. El resultat és una llista amb tants elements com esdeveniments tinguem més un.

```
library(dplyr)
library(purrr)

individual_02 <- repartir(obj = individual, events = the_events,
accumulate = TRUE)

individual_to_df <- function(individual){

df1 <- data.frame(
```



```

    birth_date = individual$birth_date,
    death_date = individual$death_date
  )

  df2<- as.data.frame(individual$anthropometric_measures)

  return(cbind(df1, df2))
}

individual_02 %>%
  purrr::map(individual_to_df) %>%
  bind_rows()

birth_date death_date weight height body_mass_index
1      <NA>      <NA>      NA     NA          NA
2    1936-11-09 <NA>      NA     NA          NA
3    1936-11-09 <NA>      65    1.72        21.97134
4    1936-11-09 <NA>      70    1.72        23.66144
5    1936-11-09 2018-05-11 70    1.72        23.66144

```

4.2.7 Avantatges de l'ús d'EDA

En aquesta secció es descriu com utilitzar les funcions i estructures implementades en `eventr` per aprofitar els avantatges d'una EDA.

4.2.7.1 *Obtenir el registre d'auditoria*

En la implementació realitzada, el registre d'auditoria és una llista d'esdeveniments (`event_list`) que emmagatzema tots els esdeveniments que han tingut lloc sobre un objecte d'interès (`the_events` en l'exemple anterior).

Per defecte, un `event_list` està estructurada com una llista. Potser és d'interès transformar aquesta estructura en un `data.frame`. Per a això es recomana transformar les capçaleres (`headers`) i els cossos (`body`) per separat. Atès que els cossos dependran de l'estructura dels esdeveniments es recomana crear una funció auxiliar (`aux_foo`) per transformar aquests cossos a un `data.frame` tal com es mostra a continuació:

```

# Headers to data.frame
the_headers <- get_header(the_events) %>%
  lapply(as.data.frame) %>%
  bind_rows()

# Auxiliar function
aux_foo = function(x) as.data.frame(x)

# bodies to data.frame
the_bodies <- get_body(the_events) %>%
  lapply(aux_foo) %>%
  bind_rows()

audit_trail <- bind_cols(the_headers, the_bodies)
audit_trail

```

	id	type	id	time
1	first-event	BIRTH		1936-11-09
2	second-id	ADD_ANTHROPOMETRIC_MEASURES		1951-01-26
3	third-id	ADD_ANTHROPOMETRIC_MEASURES		1982-05-03
4	fourth-id	DEATH		2018-05-11

	birth_date	weight	height	death_date
1	1936-11-09	NA	NA	<NA>
2	<NA>	65	1.72	<NA>
3	<NA>	70	1.72	<NA>
4	<NA>	NA	NA	2018-05-11

4.2.7.2 Testejar els esdeveniments

Suposem que volem avaluar el correcte funcionament del manipulador `anthropometric_measures_handler`. Per això crearem un nou individual amb els valors que han de prendre els seus atributs si el manipulador a avaluar funciona adequadament.

```

R> expected_individual <- new_individual()
R> expected_individual$anthropometric_measures$weight <- 65
R> expected_individual$anthropometric_measures$height <- 1.72

```

```
R> expected_individual$anthropometric_measures$body_mass_index <- 65 / 1.72^2
```

Posteriorment es crea una ocurrència amb els valors de prova (`test_ocurr`).

```
R> test_ocurr <- add_anthropometric_measures_spec(  
R+   event_id = 'second-id',  
R+   time = as.Date('1951-01-26'),  
R+   weight = 65,  
R+   height = 1.72  
R+ )
```

Mitjançant la funció `get_fun()` podem extreure la funció del manipulador que per defecte ha de tenir dos paràmetres `obj` i `event`.

```
R> foo <- get_fun(anthropometric_measures_handler)  
R>  
R> test_individual <- foo(new_individual(), test_ocurr)
```

Si tot funciona segons l'esperat, els objectes `expected_individual` i `test_individual` han de ser idèntics. En conseqüència, la funció `identical()` hauria de tornar el valor `TRUE`.

```
R> identical(test_individual, expected_individual)  
[1] TRUE
```

4.2.7.3 *Esdeveniments retroactius*

Els esdeveniments retroactius són esdeveniments que tenen efecte sobre un moment temporal anterior. Aquesta propietat és especialment útil per a recalculat l'estat d'un objecte quan ha existit algun tipus d'error en la introducció de la informació. D'aquesta manera no hi ha necessitat de preocupar-se pels efectes col·laterals que això haurà tingut sobre l'estat final. Suposem que un usuari introdueix l'altura d'un individu en centímetres i no en metres (que és la unitat per defecte).

```
correct_birth_date <- add_birth_spec(  
  event_id = 'first-event',  
  birth_date = as.Date('1936-11-09')  
)  
  
incorrect_event <- add_anthropometric_measures_spec(  
  event_id = 'second-id',  
  time = as.Date('1951-01-26'),  
  weight = 65,  
  height = 172 # Incorrect measurement  
)  
  
incorrect_events_list <- event_list(correct_birth_date,  
incorrect_event)  
  
incorrect_individual <-  
  repartir(  
    obj = new_individual(),  
    events = incorrect_events_list,  
    accumulate = FALSE  
  )  
  
individual_to_df(incorrect_individual)  
  
  birth_date  death_date  weight  height  body_mass_index  
1 1936-11-09  NA         65     172     0.002197134
```

Com s'observa en el resultat anterior l'atribut `body_mass_index` no s'ha calculat correctament. Canviant l'esdeveniment incorrecte per l'esdeveniment amb els valors correctes és possible calcular una altra vegada l'estat de l'individu.

```
correct_event <- add_anthropometric_measures_spec(  
  event_id = 'second-id',  
  time = as.Date('1951-01-26'),  
  weight = 65,  
  height = 1.72 # Correcte measurement  
)  
  
correct_events_list <- incorrect_events_list  
correct_events_list[[2]] <- correct_event  
  
correct_individual <-  
  repartir(  
    obj = new_individual(),  
    events = correct_events_list,  
    accumulate = FALSE  
  )  
  
individual_to_df(correct_individual)  
  
  birth_date  death_date  weight  height  body_mass_index  
1 1936-11-09   NA         65     1.72    21.97134
```

4.2.7.4 Estats alternatius

Els estats alternatius permeten obtenir l'estat d'un individu en moments temporals diferents o obtenir l'estat d'un individu usant dos repartidors diferents. En aquest cas es mostra com obtenir un individu en un moment temporal anterior o igual al 27 de gener de 1951. A partir de la llista d'esdeveniments de l'objecte `the_events`.

```
the_events_time <-  
  lapply(the_events, get_time) %>%
```

```
unlist()

alternative_events_list <-
  the_events[the_events_time <= as.Date('1951-01-27')]

alter_individual <- repartir(
  obj = new_individual(),
  events = alternative_events_list,
  accumulate = FALSE
) %>%
  individual_to_df()

alter_individual

  birth_date  death_date  weight  height  body_mass_index
1 1936-11-09      NA         65     1.72     21.97134
```

4.2.8 Aplicació a un cas real

En aquesta secció es mostra com podem fer ús de l'arquitectura proposada per estructurar informació mèdica procedent del registre electrònic de salut de Catalunya (Espanya). Les dades contenen informació de l'Alt Empordà, una unitat territorial situada a nord de Catalunya, a la frontera amb França i van ser obtinguts del Programa d'analítica de dades per a la recerca i la innovació en salut (PADRIS), de l'Agència de Qualitat i Avaluació Sanitàries de Catalunya (AQuAS) Govern de Catalunya.

El conjunt de dades utilitzat conté informació sobre persones que pateixen diabetis mellitus tipus 2 (DM2) i, si fos aquest el cas, la data de diagnòstic d'aquesta. A més, conté informació com l'any de naixement, si la persona està viva o no, la data de defunció (en cas que morís) i el sexe de l'individu. El conjunt de dades té un total de 66.292 individus, dels quals 32.802 són homes i 33.490 són dones. El període d'estudi va des de l'1 de gener de 2008 fins al 31 de desembre de 2018. Del conjunt total d'individus, 4.395 van patir DM2 en el període d'estudi. A la fi del període d'estudi van morir 3.391 individus. Les dades

estan disponibles en el paquet de R i podem accedir-hi mitjançant la funció `data()`.

```
data (Alt_Empordà)
head (Alt_Empordà)

  id sex dm2 dm2_date birth_year is_dead death_date
1 26000153 0 1 2002.12.05 1945 0 <NA>
2 27288814 0 1 2010-04-07 1936 0 <NA>
3 27161071 1 1 2010-08-09 1934 0 <NA>
4 25517840 0 1 2010-03-16 1938 1 2019.02.01
5 22650869 1 1 2010-11-19 1943 0 <NA>
6 27455596 1 1 2013-01-18 1938 0 <NA>
```

A continuació, es mostra com transformar les dades procedents de la taula anterior a una estructura dirigida per esdeveniments.

4.2.8.1 Estructura de l'individu

Per a aquest exemple es defineix un individu com una llista de quatre atributs: un identificador (`id`), la variable sexe (`sex`) que pot prendre els valors 0 (home) o 1 (dona), `is_dead` indica si l'individu és mort (1) o no (0) i `dm2` si l'individu en qüestió pateix o no diabetis mellitus tipus 2.

```
library(eventr)

new_individual <- function() {

  return(
    list(
      birth_date = NA,
      death_date = NA,
      anthropometric_measures = list(
        weight = NA,
        height = NA,
        body_mass_index = NA
      )
    )
  )
}
```

```
)  
  
}  
  
individual <- new_individual()
```

4.2.8.2 Crear les especificacions dels esdeveniments

Es defineixen tres especificacions: Naixement (`birth_spec`), Defunció (`death_spec`) i Diabetis mellitus tipus 2 (`dm2_spec`).

L'especificació de l'esdeveniment Naixement (`birth_spec`) té dos paràmetres d'entrada. El paràmetre `birth_year` indica en què any ha nascut l'individu. El paràmetre `sex` indica el sexe de l'individu que pot prendre dos valors: home (0) o dona (1).

```
birth_spec <- function(birth_year, sex ){  
  return(  
    event(  
      id = uuid::UUIDgenerate(),  
      type = 'BIRTH',  
      time = as.character(birth_year),  
      is_dead = FALSE,  
      sex = sex  
    )  
  )  
}
```

L'especificació de l'esdeveniment Defunció (`death_spec`) genera un identificador de l'esdeveniment mitjançant la funció `UUIDgenerate()`, assigna el tipus d'esdeveniment `DEATH` i el temps de l'esdeveniment com la data de la defunció `death_date`. Per defecte, al paràmetre `is_dead` se li assigna el valor `TRUE` quan es fa ús d'aquesta especificació.


```

death_spec <- function(death_date){
  return(
    event(
      id = uuid::UUIDgenerate(),
      type = 'DEATH',
      time = as.character(death_date),
      is_dead = TRUE
    )
  )
}

```

L'especificació de l'esdeveniment diabetis mellitus tipus 2 assigna el tipus `DIABETS_MELLITUS_TYPE_2` a l'esdeveniment i assigna el valor de l'atribut `dm2` a `TRUE` per defecte.

```

dm2_spec <- function(event_time) {

  return(
    event(
      id = uuid::UUIDgenerate(),
      type = 'DIABETES_MELLITUS_TYPE_2',
      time = as.character(event_time),
      dm2 = TRUE
    )
  )
}

```

Un cop definides les especificacions s'implementen els manipuladors.

4.2.8.3 *Crear els manipuladors*

Els manipuladors s'encarreguen de transformar els paràmetres d'un esdeveniment en els atributs de l'individu.

El manipulador `birth_handler` és l'encarregat de manipular els esdeveniments de tipus `BIRTH`. A partir dels esdeveniments de tipus `BIRTH` assigna els atributs `is_dead` i `sex`.

```
birth_handler <- handler(  
  type = 'BIRTH',  
  FUN = function(obj, event){  
    obj$is_dead <- get_body_attr(event, 'is_dead')[[1]]  
    obj$sex <- get_body_attr(event, 'sex')[[1]]  
    return(obj)  
  })
```

El manipulador `death_handler` és l'encarregat de manipular els esdeveniments de tipus `DEATH`. Modifica l'atribut `is_dead` de l'individu.

```
death_handler <- handler(  
  type = 'DEATH',  
  FUN = function(obj, event){  
    obj$is_dead <- get_body_attr(event, 'is_dead')[[1]]  
    return(obj)  
  })
```

El manipulador `dm2_handler` és l'encarregat de manipular els esdeveniments de tipus `DIABETES_MELLITUS_TYPE_2`. Modifica l'atribut `dm2` de l'individu.

```
dm2_handler <- handler(  
  type = 'DIABETES_MELLITUS_TYPE_2',  
  FUN = function(obj, event){  
    obj$dm2 <- get_body_attr(event, 'dm2')  
    return(obj)  
  })
```

4.2.8.4 Crear el repartidor

Un cop creats els manipuladors es crea el nou repartidor mitjançant la funció `dispatcher()`.

```
the_dispatcher <- dispatcher(  
  handlers = handlers_list(  
    birth_handler,  
    death_handler,  
    dm2_handler  
  )  
)
```

4.2.8.5 Crear les ocurrències dels individus

A continuació, es crea una llista amb els esdeveniments per a cada individu. Cada element de la llista individuals representa un individu del conjunt de dades original i està format per una llista d'esdeveniments (`event_list`).

```
> # split data.frame by individual  
> individuals <- split(alt_emporda, f = alt_emporda$id)  
>  
> head(individuals, 3)  
$`12602285`  
      id sex dm2 dm2_date birth_year is_dead death_date  
24002 12602285 1 0 <NA>      1971      0 <NA>  
  
$`12602295`  
      id sex dm2 dm2_date birth_year is_dead death_date  
24003 12602295 0 0 <NA>      1992      0 <NA>  
  
$`12622895`  
      id sex dm2 dm2_date birth_year is_dead death_date  
24004 12622895 0 0 <NA>      2018      0 <NA>
```

Es creen els esdeveniments per a cada un dels elements de la llista individuals.

```
individuals <- lapply(individuals, function(x){

# birth event

the_events <- birth_spec(
  birth_year = unique(x$birth_year),
  sex = ifelse(
    unique(x$sex) == 1,
    'woman', ifelse(
      unique(x$sex) == 0,
      'man',
      'unknown'
    )
  )
)

# create death event if is_dead == 1

if(x$is_dead == 1){

  the_events <- the_events +
  death_spec(death_date = unique(x$death_date)[[1]])

}

# create diabetes mellitus type 2 event if dm2 == 1

if(x$dm2 == 1){

  the_events <- the_events +
  dm2_spec(
    event_time = unique(x$dm2_date)
  )

}

if(class(the_events)[1] == "event") the_events <-
event_list(the_events)

return(the_events)

}
```

```
}))
```

Un cop creats els esdeveniments dels individus ja podem fer ús de les funcions de eventr

4.2.8.6 Executar els esdeveniments per a cada individu

```
repartir <- get_dispatch(the_dispatcher)
ids <- names(individuals)

li <- list()

for(id in ids){

  x <- individuals[id]
  li[[id]] <- repartir(
    obj = new_individual(id = names(x)),
    events = x[[1]],
    accumulate = FALSE
  ) %>% as.data.frame()

}
```

```
li %>% bind_rows() %>% head(6)
```

	id	sex	is_dead	dm2
1	12602285	woman	FALSE	FALSE
2	12602295	man	FALSE	FALSE
3	12622895	man	FALSE	FALSE
4	12622896	woman	FALSE	FALSE
5	12622899	woman	FALSE	FALSE
6	12623072	woman	FALSE	FALSE

En l'exemple anterior s'ha fet servir el valor `accumulate = FALSE`, per tant, el resultat final és una única fila per individu. Si ens interessa obtenir una fila per a cada esdeveniment que ha tingut lloc podem usar el paràmetre `accumulate = TRUE` tal com es mostra a continuació

```
second_list <- list()

for(id in ids){

  x <- individuals[id]
  second_list[[id]] <- repartir(
    obj = new_individual(id = names(x)),
    events = x[[1]],
    accumulate = TRUE
  ) %>%
  purrr::map(as.data.frame) %>%
  bind_rows() %>%
  # Elim first row wich contains default individual
  .[-1,]

}

second_list <- second_list %>% bind_rows()
second_list[25:31,]
```

	id	sex	is_dead	dm2
25	12752496	man	FALSE	FALSE
26	12752497	man	FALSE	FALSE
27	12752498	woman	FALSE	FALSE
28	12752576	man	FALSE	FALSE
29	12752583	man	FALSE	FALSE
30	12752583	man	TRUE	FALSE
31	12752638	man	FALSE	FALSE

L'usuari amb identificador 12.752.583 té dues files atès que a la fi de l'estudi era mort.

4.2.8.7 *Obtenir l'Audit Trail*

El registre d'auditoria permet revisar tots els esdeveniments que han tingut lloc per a un individu i quins atributs s'han vist modificats en cada un dels esdeveniments. Per això és una eina útil per identificar falles de codi. A continuació, es mostra com obtenir el registre d'auditoria per al conjunt de dades individuals.

```
audit_trail <- lapply(individuals, function(occurrences){

  if(length(occurrences) == 1){

    headers <-
      get_header(occurrences) %>%
      as.data.frame() %>%
      rename(event_id = id)
    bodies <-
      get_body(occurrences) %>%
      as.data.frame()

    return(bind_cols(headers, bodies))

  } else {

    if(length(occurrences) > 1){

      headers <- lapply(occurrences, function(x) {
        get_header(x) %>%
          as.data.frame() %>%
          rename(event_id = id)
      }) %>% bind_rows()

      bodies <- lapply(
        X = occurrences,
        FUN = function(x) {
          get_body(x) %>% as.data.frame()
        }) %>% bind_rows()

      return(bind_cols(headers, bodies))

    }

  }

}
```

```

}

})

audit_trail %>%
  bind_rows(.id = 'individual_id') %>%
  head()

  individual_id          event_id type time is_dead sex dm2
1      12602285 94624a1b-4dad-4366-9a3e-804c7fc82b4e BIRTH 1971 FALSE
woman NA
2      12602295 6dbdd9aa-dfcb-4a77-bfe3-7fbbb7bf206f BIRTH 1992 FALSE
man NA
3      12622895 2bbc68f3-b055-4174-8f62-16ffc8cc4f10 BIRTH 2018 FALSE
man NA
4      12622896 2266137f-703c-4d02-ab13-166543ee71f7 BIRTH 2018 FALSE
woman NA
5      12622899 48828e85-d8eb-46ec-bca6-09bc5d90e627 BIRTH 2018 FALSE
woman NA
6      12623072 8b493907-e2de-47e8-b02f-91dce32f4fe8 BIRTH 2009 FALSE
woman NA

```

Com es mostra en el codi anterior l'auditoria proporcionada per l'individu mostra, l'identificador de l'esdeveniment (`event_id`), el tipus d'esdeveniment que ha tingut lloc (`type`), el moment en què aquest esdeveniment ha tingut lloc (`time`) i els valors dels atributs en cada un dels esdeveniments. Per a l'exemple mostrat, els atributs modificats són `is_dead` i `sex` donat que els esdeveniments que es mostren com a resultat de la funció `head` pertanyen tots a esdeveniments de tipus `BIRTH`.

Per avaluar l'impacte de la llibreria **eventr** s'ha fet ús del nombre de descarregues realitzades des del CRAN (Institute for Statistics and Mathematics of WU (Wirtschaftsuniversität Wien), 2021). Per obtenir el nombre de descàrregues s'ha fet servir la llibreria `cranlogs` (Csárdi, 2021). En el període comprès entre el dia 26 de Maig de 2020 (quan es va pujar al repositori del CRAN) i l'actualitat el paquet `eventr` s'ha descarregat 3415 cops. El que significa que en mitjana es descarrega 293 cops al mes o 11 cops al dia. La Taula 10 mostra el nombre de descàrregues mensuals.

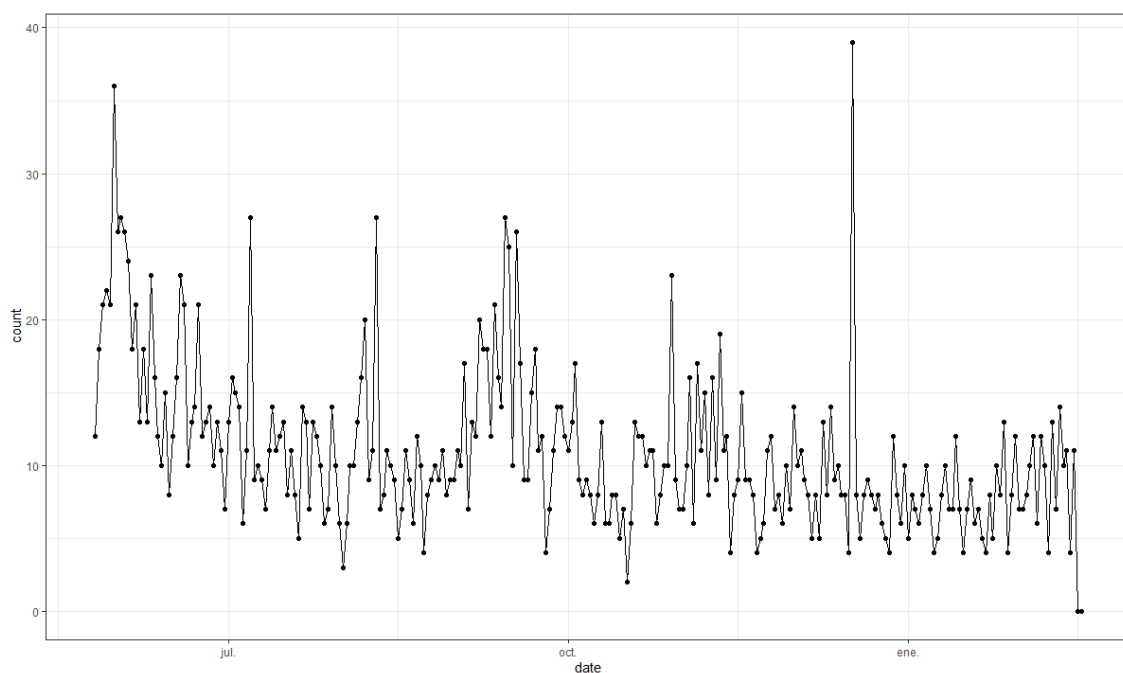


Figura 21 Descàrregues de la llibreria eventr des del 2020-05-26 fins al 2021-06-04. Font: Elaboració pròpia a partir de dades obtingudes mitjançant la llibreria cranlogs.

Taula 10 Nombre de descàrregues mensuals del paquet eventr des del repositori del CRAN. Font: Elaboració pròpia a partir de les dades obtingudes mitjançant el paquet cranlogs..

Any	Mes	N
2020	Maig	130
2020	Juny	480
2020	Juliol	342
2020	Agost	307
2020	Setembre	430
2020	Octubre	292

2020	Novembre	295
2020	Desembre	287
2021	Gener	228
2021	Febrer	293
2021	Març	322

Font: Elaboració pròpia.

4.3 REPOSITORI DE DADES INDIKA

El repositori de dades d'Indika, allotjat a la pàgina web de GitHub es pot trobar a l'enllaç <https://github.com/alvarofranq/indika-data-repository>. La Figura 22 mostra la pàgina principal del repositori. El repositori compta amb 2 en el moment d'escriure la tesi.

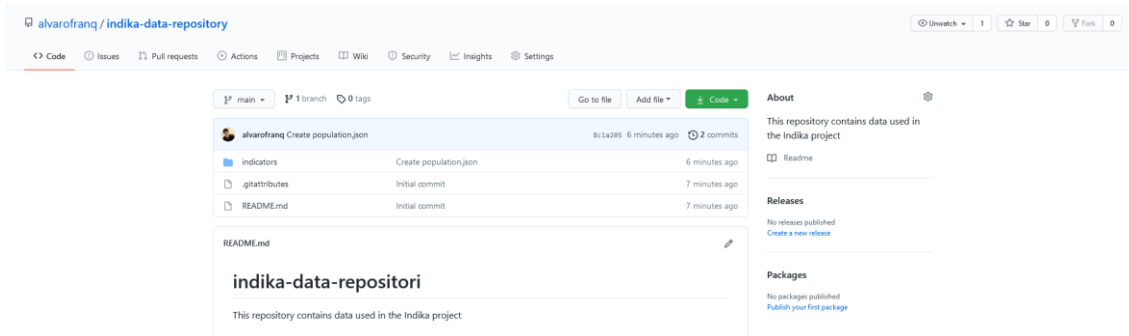


Figura 22 Pàgina inicial del repositori de dades d'Indika. Font: Elaboració pròpia.

És possible accedir a la informació dels indicadors mitjançant comandes de R i la llibreria *jsonlite* tal com es mostra a continuació:

```
> URL <- 'https://raw.githubusercontent.com/alvarofranq/indika-data-repository/main/indicators/population.json'

> indicator <- jsonlite::read_json(URL, simplifyVector = TRUE)
```

```
> head(indicator$data)
  municipalityCode      municipalityName mens womens total
1          170010          Agullana    431    432    863
2          170031          Albanyà     79     76    155
3          170118      Armentera, l'  478    473    951
4          170123 Avinyonet de Puigventós 818    823   1641
5          170160          Bàscara    521    480   1001
6          172348           Biure     125    116    241
```

En l'exemple anterior a partir de la URL es descarrega la informació corresponent a l'indicador *population* que conté el nombre d'homes i dones a cada municipi de l'Alt Empordà.

De forma similar podem fer ús de les llistes d'indicadors per descarregar múltiples indicadors. A continuació és mostra un exemple de com fer servir R per descarregar els indicadors inclosos en una llista d'indicadors, en aquest cas la llista *peopleAndPopulation*.

```
library(jsonlite)
library(tidyverse)
library(rvest)

> URL <- 'https://raw.githubusercontent.com/alvarofranq/indika-data-
repository/main/indicators-list/peopleAndPopulation.json'
indicators_list <- jsonlite::read_json(URL, simplifyVector = TRUE)

> people_and_population <- lapply(indicators_list,
jsonlite::read_json, simplifyVector = TRUE)

> str(people_and_population, max.level = 2)
List of 2
 $ :List of 2
  ..$ meta:List of 6
  ..$ data:'data.frame':    68 obs. of  5 variables:
 $ :List of 2
  ..$ meta:List of 6
  ..$ data:'data.frame':   126 obs. of  5 variables:
```

5 DISCUSSIÓ

El propòsit d'aquesta tesi és desenvolupar des d'un punt de vista científic un observatori de salut i social a l'Alt Empordà i dotar-lo de les eines necessàries per al seu funcionament inicial i futur. Les següents seccions proporcionen una discussió global dels principals resultats, juntament amb les seves limitacions, punts forts i futures orientacions de recerca. A continuació s'inicia el procés de discussió partint de tres punts fonamentals. Primerament la construcció de l'observatori Indika, seguidament, la construcció del repositori públic d'informació i finalment els punts forts i febles de fer ús d'una arquitectura basada en esdeveniments per estructurar la informació i la seva implementació a la llibreria *eventr*.

5.1 INDIKA POL DE SALUT I SOCIAL DE L'ALT EMPORDÀ

A partir del treball de Caiaffa *et al.* (2014) com a marc de referència podem fer una avaluació de les característiques de l'observatori i veure si aquestes s'ajusten al concepte d'observatori creat. A continuació es descriu punt per punt la valoració de la feina realitzada. L'avaluació es pot estructurar en quatre parts: la missió de l'observatori, la seva governança, la integració de la informació territorial i l'evidència generada.

L'observatori Indika té una missió clara i aquesta va alineada amb els objectius de desenvolupament sostenible. Té la capacitat de generar coneixement sobre temes d'interès comarcal i afavorir la formació en els diferents àmbits. És un observatori multidisciplinar, el que aporta diferents visions al tractament de les dades i la generació de coneixement. A més a més és capaç de produir material per a la presa de decisions com "Indicadors Clau de l'Alt Empordà" (Franquet *et al.*, 2020).

La governança, entesa com la manera d'organitzar l'observatori i la forma en que aquest és relaciona amb d'altres entitats, no està clara, així com tampoc les responsabilitats dels diferents integrants d'aquest. Cal especificar millor qui és l'encarregat de cada una de les accions i de les decisions que es porten a terme així com les entitats que en formen part i com en formen part (aportant capital

Discussió

econòmic o capital humà). Aquest és el cas de l'Ajuntament de Figueres o del consell comarcal que en formen part, però no queda clar què aporten al mateix ni quines responsabilitats tenen sobre la informació que en surt publicada. Un altre punt important en la governança és aconseguir diversificar l'obtenció de capital per evitar que tot depengui d'una única entitat per tal d'assegurar la sostenibilitat a llarg termini d'Indika. En el moment de lliurar aquesta tesi, Indika ha obtingut finançament públic pel tal de poder continuar amb la seva activitat en el marc del projecte PECT Innovació Social-EMISSE

Indika està dotada d'una bona capacitat d'intel·ligència de negoci el que li permet identificar recursos d'interès públics i privats i fer-ne ús. L'observatori disposa d'un repositori d'informació replicable per a altres comarques i ampliable que facilita la traçabilitat de les anàlisis elaborades i permet la integració amb d'altres projectes que s'estan portant a terme a nivell comarcal com és l'Observatori de Desigualtats Socials i de Salut de Dipsalut. La reproductibilitat de la informació publicada també està garantida atès que totes les anàlisis han estat realitzades mitjançant el llenguatge de programació R (R Core Team, 2020) què es obert i accessible per tothom. Els diferents codis emprats es poden demanar directament als responsables de l'observatori. Amb l'aplicació de la COVID-19 Indika ha demostrat tenir la capacitat de monitorar l'estat del territori de forma efectiva i les diferents aparicions a la premsa mostren que també té la capacitat de fer difusió de les accions realitzades.

En el procés de desenvolupament de l'observatori s'ha fet servir una metodologia de treball àgil (Fowler i Highsmith, 2001). Les metodologies de treball àgil molt emprades en el desenvolupament de software posen el focus en el disseny iteratiu del producte on els requisits i solucions evolucionen amb el temps segons les necessitats del projecte. Els equips de treball que han de ser autoorganitzats i multidisciplinars, com és el cas d'Indika, tenen com a objectiu resoldre de manera col·laborativa i autoorganitzada les diferents problemàtiques que van sorgint al llarg del procés.

En el cas d'Indika aquest procés ha consistit en reunions periòdiques amb l'equip de l'observatori. En aquestes reunions es decideixen els objectius a assolir a curt

Discussió

termini, com per exemple, la creació de la pàgina web o les diferents infografies i es prioritzen les accions a desenvolupar, es divideixen les tasques i s'assigna el temps per portar a terme aquestes accions. Aquesta periodització d'accions és coneguda com a *sprints*.

Les reunions s'estructuren en tres parts fonamentals: posada en comú dels resultats obtinguts i estat de les accions desenvolupades durant aquell període, i en cas de sorgir alguna problemàtica es debaten accions concretes per solucionar-la. A partir d'aquesta discussió s'ordenen les prioritats i finalment es decideixen les accions del nou *sprint*.

Aquest procés inclou també la realització de reunions per avaluar de forma retrospectiva quines són aquelles coses que han funcionat bé, quines s'han de millorar i quines no han funcionat. D'aquesta manera es pretén millorar els punts febles i potenciar els punts forts de l'equip.

La pandèmia ha fet que canviïn les prioritats informatives de l'observatori. En relació a la pandèmia s'han dut a terme dues accions principals. La creació d'una aplicació web on poder veure quin és l'estat de la comarca i una infografia resum passat un any del primer cas de COVID-19 confirmat a la comarca. D'aquestes hem obtingut els aprenentatges que esmentem a continuació.

La creació d'aplicacions web específiques, com és el cas de l'aplicació esmentada en la secció 4.1.2, té un cost de recursos massa elevat per a l'ús real que s'ha fet d'aquesta. Tot i haver estat una aplicació útil a l'inici de la pandèmia no ho ha estat tant al final, on podem trobar moltes altres opcions que ofereixen informació més detallada. Donat que les dades van canviant en funció de les noves descobertes i les necessitats del moment, per poder oferir una aplicació actualitzada i que aportí valor cal tenir una persona dedicada a aquesta tasca. La rentabilitat d'aquest tipus d'accions ha de ser avaluada tenint en compte molts factors en un observatori com Indika.

Discussió

Per contra, la creació d'un informe en clau de resum anual descrit a la secció 4.1.1.3 sí que ha estat una bona aposta, ja que l'impacte en els mitjans ha estat elevat permetent donar a conèixer l'observatori i la situació comarcal.

Cal destacar que gràcies a la metodologia de treball emprada ha sigut fàcil adaptar-se a les noves necessitats que anaven sorgint com a conseqüència de la pandèmia.

5.2 REPOSITORI DE DADES INDIKA

El repositori de dades Indika conté informació rellevant del territori en format JSON que pot ser descarregada mitjançant la pàgina GitHub (2020) o bé fent ús de les URL als diferents indicadors directament tal com es mostra a l'apartat 4.3. L'ús de URLs facilita l'explotació de les dades mitjançant llenguatges de programació com pot ser R (R Core Team, 2020). Gràcies a les llistes d'indicadors es possible descarregar múltiples indicadors de manera simultània.

Fer ús de la plataforma GitHub (2020) permet als usuaris accedir a la informació i participar de forma activa en la seva actualització i recollida. A més hi ha la opció de descarregar tota la informació per fer-ne ús de manera local. Qualsevol entitat o persona és capaç d'incloure nous indicadors. Només ha d'assegurar-se que l'estructura segueix les característiques descrites en la secció 3.2.

Aquest punt planteja la necessitat de crear mecanismes de verificació de la informació. Aquest mecanismes poden ser: la revisió de les dades en cas que aquestes siguin de domini públic o bé la necessitat de posar-se en contacte amb Indika primer, en cas de tractar-se de dades procedents d'entitats o persones privades, de manera que es pugui verificar la seva autenticitat.

Donades les característiques del repositori és possible saber quina persona o entitat ha fet cada actualització atorgant a cadascú el mèrit que mereix. Donat, a més a més, que aquest és obert i accessible per qualsevol persona podem dir que el repositori de dades Indika compleix amb els requisits definits en l'apartat 3.2.

5.3 EVENTR

La llibreria *eventr* ha estat publicada al *Comprehensive R Archive Network (CRAN)* (Institute for Statistics and Mathematics of WU (Wirtschaftsuniversität Wien), 2021) què és el repositori de referència per a totes les llibreries de R. Per tal que la llibreria es publiqui al CRAN s'ha de passar una revisió duta a terme per un expert en el tema, el que significa que la qualitat de la mateixa és prou bona per poder ser publicada. La publicació al CRAN implica que l'estructura d'*eventr* compleix els requisits de la comunitat de desenvolupador donant validesa a la mateixa. Per tant, la llibreria *eventr* està dotada d'una documentació adequada que permet entendre quina és la seva utilitat i com es pot fer-ne ús.

A més a més la llibreria està publicada a la pàgina web de codi obert GitHub i disponible a través d'aquest enllaç: <https://github.com/alvarofranq/eventr>. A partir d'aquesta pàgina web és possible notificar possibles problemàtiques derivades de l'ús d'*eventr* i afegir o modificar, de forma col·laborativa, les seves funcionalitats.

El procés de creació de *eventr* s'ha estructurat en tres passos: Identificació de les necessitats, disseny de la implementació i implementació en R. La identificació dels requeriments s'ha fet de forma adequada centrant la problemàtica en la necessitat de com cal estructurar la informació per poder reutilitzar-la mitjançant diverses eines d'anàlisi estadístic.

Respecte al disseny realitzat s'ha centrat en una arquitectura basada en esdeveniments bàsica, això dona peu a múltiples actualitzacions de la mateixa incrementant la seva complexitat o construint altres llibreries a sobre que aporten més riquesa i utilitats.

Tal com ja s'ha comentat la implementació realitzada compleix tots els criteris definits per la comunitat de desenvolupadors així com la documentació necessària per a la seva utilització com són els exemples amb dades reals.

Una de les problemàtiques de la implementació de la llibreria és la seva alta complexitat derivada de la pròpia dificultat tècnica de implantació d'arquitectures basades en esdeveniments en entorns reals. La transformació de dades

Discussió

estructurades mitjançant altres arquitectures pot ser complexa desincentivant l'ús d'aquest tipus de metodologia.

6 CONCLUSIONS I TREBALL FUTUR

A partir dels resultats obtinguts en aquesta tesis es pot concloure que:

- Els observatoris de salut d'àmbit territorial – com pot ser una comarca - són una bona eina per dotar als territoris de la intel·ligència de negoci en salut necessària per a la presa de decisions. És necessària però una col·laboració a un nivell superior per a no duplicar esforços i compartir tecnologies així com eines d'anàlisi amb altres elements existents. A més la funció d'un observatori de salut comarcal ha de centrar-se sobretot en la creació de xarxes de col·laboració efectives que a l'hora generen coneixement científic sobre les accions realitzades en el territori com a resultat de l'anàlisi de la informació disponible.
- L'arquitectura basada en esdeveniments facilita el tractament estadístic des d'un punt de vista científic de la informació disponible en els registres electrònics mèdics entre d'altres. La seva implementació és costosa en temps i capital humà. Cal tenir en compte aquests dos factors a l'hora de fer-ne ús i desenvolupar els processos.
- La creació d'un repositori de dades obert com el repositori de dades d'Indika ha d'anar acompanyat d'un procés de difusió del mateix així com de la creació de material didàctic que permetí educar als possibles usuaris en l'ús del repositori i en la forma de col·laborar. Aquest material ha de tenir presents els diferents perfils professionals existents incloent aquelles persones amb la capacitat tècnica de col·laborar directament o bé les entitats que disposen de dades, però no disposen de la capacitat per fer-ne una explotació adequada acompanyant-les en el procés d'obrir informació.

A partir de la discussió realitzada és possible identificar alguns punts o línies de treball i de recerca futures com són:

Conclusions i treball futur

- Dissenyar una estratègia comunicativa que faci ús de les xarxes socials per fer difusió de la feina realitzada per Indika. Especialment si es vol arribar a la població més jove del territori.
- Concretar millor la implicació de les diferents aliances territorials emmarcant aquesta implicació en accions concretes que tinguin un impacte directe sobre l'Alt Empordà.
- Dissenyar una estratègia de col·laboració amb les diferents entitats territorials. És a dir, redefinir i concretar com es pot treballar de forma conjunta amb altres agents territorials i què considerem que és una col·laboració efectiva.
- Avançar en una cooperació i treball escalable amb altres observatoris d'àmbit territorial o d'abast superior.
- Dissenyar una estratègia a llarg termini on s'identifiquin els problemes més importants del territori i s'especifiquen de manera clara i concisa els objectius de l'observatori com a agent territorial emmarcant aquestes accions amb els objectius de desenvolupament sostenible.

7 REFERÈNCIES

Adobe Inc. (01 de 04 de 2019). Adobe Illustrator. Obtenido de <https://adobe.com/products/illustrator>

Asamblea General de las Naciones Unidas. (2015). *Transformar nuestro mundo: la Agenda 2030 para el Desarrollo Sostenible*. Obtenido de Transformar nuestro mundo: la Agenda 2030 para el Desarrollo Sostenible: https://www.un.org/ga/search/view_doc.asp?symbol=A%2F70%2FL.1

Boukraâ, D., & Boussaid, O. (09 de 2009). *Complex Data Warehouse Modelling*. Recuperado el 24 de 03 de 2021, de https://www.researchgate.net/profile/Doukifli-Boukraa/publication/228574017_Complex_Data_Warehouse_Modelling/links/549d886e0cf2b803713a787e/Complex-Data-Warehouse-Modelling.pdf

Caiaffa, W. T. (2014). Developing a conceptual framework of urban health observatories toward integrating research and evidence into urban policy for health and health equity. *Journal of urban health : bulletin of the New York Academy of Medicine*, 91(1), 1-16. doi:<https://doi.org/10.1007/s11524-013-9812-0>

Cairo, A. (2016). *The truthful art: Data, charts, and maps for communication*. Estats Units d'Amèrica: New Riders.

Castillo, C. (2015). Developing an academia-based public health observatory: the new global public health observatory with emphasis on urban health at Johns Hopkins Bloomberg School of Public Health. (S. P. Health, Ed.) *Cadernos de saude publica*, 31(supl.1), 286-293. doi:<https://doi.org/10.1590/0102-311X00132914>

Chacon, S., & Straub, B. (2014). *Pro git*. Nova York, Estats Units d'Amèrica: Apress, Berkeley, CA. doi:<https://doi.org/10.1007/978-1-4842-0076-6>

Referències

- Chang, W., Cheng, J., Allaire, J., Xie, Y., & McPherson, J. (20 de 11 de 2020). *shiny: Web Application Framework for R*. Obtenido de <https://CRAN.R-project.org/package=shiny>
- Cheng, J., Sievert, C., Chang, W., Xie, Y., & Allen, J. (2020). *htmltools: Tools for HTML*. Recollit de <https://github.com/rstudio/htmltools>
- Cofiño, R., Pasarín, M. I., & Segura, A. (2012). ¿Cómo abordar la dimensión colectiva de la salud de las personas? Informe SESPAS 2012. *Gaceta Sanitaria*, 26, 88-93. doi:<https://doi.org/10.1016/j.gaceta.2011.07.030>
- Comisión de Estadística de la ONU. (03-06 / 03 / 2020). *Informe sobre el 51er período de sesiones*. Recollit de unstats: <https://unstats.un.org/unsd/statcom/51st-session/documents/2020-37-FinalReport-S.pdf>
- Comisión de Estadística de las Naciones Unidas. (7-10 / 03 / 2017). Consultat el 19 / 03 / 2021, a <https://undocs.org/pdf?symbol=es/E/2017/24>
- Comisión para Reducir las Desigualdades Sociales en Salud en España. (2012). Propuesta de políticas e intervenciones para reducir las desigualdades sociales en salud en España. *Gaceta Sanitaria*, 26(2), 182-189. doi:10.1016/j.gaceta.2011.07.024
- Creative Commons. (s.f.). CC0. Recuperado el 16 de 02 de 2020, de Creative Commons: <https://creativecommons.org/share-your-work/public-domain/cc0>
- Csárdi, G. (2021). *cranlogs: Download Logs from the 'RStudio' 'CRAN' Mirror*. R package version 2.1.1. Obtenido de <https://r-hub.github.io/cranlogs>
- Dahlgren, G., & Whitehead, M. (01 / 1991). Policies and strategies to promote social equity in health. *Institute for Futures Studies, Arbetsrapport*, 14. Recollit de https://www.researchgate.net/profile/Goeran-Dahlgren/publication/5095964_Policies_and_strategies_to_promote_social_equity_in_health_Background_document_to_WHO_-

Referències

[_Strategy_paper_for_Europe/links/569540f808aeab58a9a4d946/Policies-and-strategies-to-promote-so](#)

Dahlgren, G., & Whitehead, M. (2006). *European strategies for tackling social inequities in health: Levelling up Part 2*. WHO Regional office for Europe Copenhagen. Recollit de https://www.euro.who.int/__data/assets/pdf_file/0018/103824/E89384.pdf

Daponte Codina, A., Cabrera-León, A., Mateo Rodríguez, I., Espinosa de los Monteros, E., Arroyo-Borrell, E., Saez, M., . . . Bolívar. (2019). *Atlas de los determinantes sociales de la salud en España 2019: evolución y variabilidad entre las Comunidades Autónomas*. Publicaciones EASP.

Demski, H., Garde, S., & Hildebrand, C. (2016). Open data models for smart health interconnected applications: the example of openEHR. *BMC Medical Informatics and Decision Making*, 16(1), 137. doi:<https://doi.org/10.1186/s12911-016-0376-2>

Departament de Salut de la Generalitat de Catalunya. (17 / 03 / 2021a). *Registre de casos de COVID-19 realitzats a Catalunya. Segregació per sexe i municipi*. Recollit de Dades Obertes Catalunya: <https://analisi.transparenciacatalunya.cat/Salut/Registre-de-casos-de-COVID-19-realitzats-a-Catalun/jj6z-iyrp>

Departament de Salut de la Generalitat de Catalunya. (16 / 03 / 2021b). *Registre de casos de COVID-19 realitzats a Catalunya. Segregació per sexe i edat*. Recollit de Dades Obertes Catalunya: <https://analisi.transparenciacatalunya.cat/Salut/Registre-de-casos-de-COVID-19-realitzats-a-Catalun/qwj8-xpvk>

Departament de Salut de la Generalitat de Catalunya. (16 / 03 / 2021c). *Registre de defuncions per COVID-19 a Catalunya. Segregació per sexe i comarca*. Recollit de Dades Obertes Catalunya: <https://analisi.transparenciacatalunya.cat/Salut/Registre-de-defuncions-per-COVID-19-a-Catalunya-Se/uqk7-bf9s>

Referències

- Departament de Salut de la Generalitat de Catalunya. (16 / 03 / 2021d). *Vacunació per al COVID-19 per comarca*. Consultat el 16 / 03 / 2021, a Dades Obertes Catalunya: <https://analisi.transparenciacatalunya.cat/Salut/Vacunaci-per-al-COVID-19-per-comarca/cuwj-bh3b>
- Devlin, H., Schenk, T. J., Leynes, G., Lucius, N., Malc, J., Silverberg, M., & Schmeideskamp, P. (2019). *RSocrata: Download or Upload 'Socrata' Data Sets*. Obtenido de <https://CRAN.R-project.org/package=RSocrata>
- ECMA International. (12 / 2017). *ECMA-404 The JSON Data Interchange Syntax*. Consultat el 19 / 03 / 2021, a ECMA International: https://www.ecma-international.org/wp-content/uploads/ECMA-404_2nd_edition_december_2017.pdf
- Erb, B., Habiger, G., & Hauck, F. J. (2016). On the Potential of Event Sourcing for Retroactive Actor-Based Programming. *First Workshop on Programming Models and Languages for Distributed Computing*(9781450347754), 1-5. doi:10.1145/2957319.2957378
- European Commission. (sense data). *European Commission*. Consultat el 12 / 01 / 2020, a The European Core Health Indicators (ECHI) shortlist of 88 health indicators identified by policy area: https://ec.europa.eu/health/sites/health/files/indicators/docs/echi_shortlist_by_policy_area_en.pdf
- EUROSTAT. (2019). *Key figures on Europe — Statistics illustrated — 2019 edition*. (U. B.—D. Helene Strandell and Pascal WolffEurostat, Ed.) Luxembourg, Luxembourg, Luxembourg: EUROSTAT. doi:10.2785/415908
- EUROSTAT. (2020). *Key figures on Europe — Statistics illustrated — 2020 edition*. Luxembourg, Luxembourg, Luxembourg: EUROSTAT. doi:10.2785/894405

Referències

Extensible Markup Language (XML) 1.0 (Fifth Edition). (26 de 11 de 2008).

Recuperado el 17 de 02 de 2020, de W3C: <https://www.w3.org/TR/xml/>

Federación española de municipios y provincias, Red de entidades locales por la transparencia y participación ciudadana, wolter kluwer. (24 de 04 de 2018). *datos.gob.es*. Obtenido de Datos abiertos: guía estratégica para su puesta en marcha y conjuntos de datos mínimos a publicar: https://datos.gob.es/sites/default/files/doc/file/guia_datos_abiertos.pdf

Few, S. (2013). *Information Dashboard Design: Displaying data for at-a-glance monitoring*. (B. Pierce, Ed.) El Dorado Hills, CA 95762: Analytics Press.

Fowler, M. (2004). *UML distilled: a brief guide to the standard object modeling language*. Estats Units d'Amèrica: Addison-Wesley Professional.

Fowler, M. (25 / 01 / 2006). *Focusing on Events*. Consultat el 24 / 03 / 2021, a martinfowler.com:
<https://www.martinfowler.com/eaDev/EventNarrative.html>

Fowler, M. (07 / 02 / 2017). *What do you mean by "Event-Driven"?* Consultat el 24 / 03 / 2021, a martinfowler.com:
<https://www.martinfowler.com/articles/201701-event-driven.html>

Franquet Bonet, A., Plaja Roman, P., Iglesias Hernandez, E., Barceló Rado, M. A., & Saez Zafra, M. (2020). *Indicadors Clau de l'Alt Empordà*. Figueres: Departament de Comunicació de la Fundació Salut Empordà. doi:doi: 10.48334/I9109-5826-6199-

Franquet, A., Barceló, M. A., Saez, M., & Plaja, P. (26 de 05 de 2020). *eventr: Create Event Based Data Architectures*. Obtenido de <https://cran.r-project.org/web/packages/eventr/index.html>

Free Software Foundation. (31 / 11 / 2020). <https://www.gnu.org/licenses/gpl-3.0.en.html>. Recollit de GNU Operating System: <https://www.gnu.org/licenses/gpl-3.0.en.html>

Referències

- Fundació Salut Empordà. (2019). *Infografies: Demografia*. Consultat el 01 / 03 / 2021, a Indika Pol de Salut i Social de l'Alt Empordà: <https://www.indikasalut.cat/wp-content/uploads/2021/01/01-poblacio-alt-emporda.pdf>
- Generalitat de Catalunya. (2018). *Guia de visualització de dades*. Barcelona: Generalitat de Catalunya. Recollit de <http://atenciociudadana.gencat.cat/web/.content/manuals/guia-visualitzacio-dades.pdf>
- Generalitat de Catalunya. (20 de 11 de 2020). *Dades obertes: Govern obert*. Obtenido de Govern obert: <https://analisi.transparenciacatalunya.cat/>
- GitHub. (30 de 11 de 2020). Obtenido de GitHub: <https://github.com/>
- Griggs, D., Stafford-Smith, M., Gaffney, O., Rockström, J., Öhman, M. C., Shyamsundar, P., . . . Noble, I. (2013). Sustainable development goals for people and planet. *Nature*, *495*, 305-307. doi:<https://doi.org/10.1038/495305a>
- Hametner, M., Kostetckaia, M., Setz, I., Bley, S.-J., Gebhard, F., Mayer, C., . . . Steurer, A. (2019). *Sustainable development in the European Union: Overview of progress towards the SDGs in an EU context*. Luxembourg: Publications Office of the European Union,. doi:10.2785/58838
- Hasenkopf, C., Flasher, J., Veerman, O., & DeWitt, H. L. (2015). OpenAQ: A Platform to Aggregate and Freely Share Global Air Quality Data. *AGU Fall Meeting Abstracts*, A31D--0097.
- Hripcsak, G., Duke, J., Shah, N., Reich, C., Huser, V., Schuemie, M., . . . Ryan, P. (2015). Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. *Studies in health technology and informatics*, *216*, 574-578.

Referències

Idescat. (2019a). *Padró municipal d'habitants. Per sexe*. Consultat el 10 / 02 / 2021, a Idescat: <https://www.idescat.cat/pub/?id=pmh&n=446&geo=com:02>

Idescat. (27 / 12 / 2019b). *Padró municipal d'habitants. Per sexe*. Consultat el 10 / 02 / 2021, a Idescat: <https://www.idescat.cat/pub/?id=pmh&n=446&geo=com%3A02&by=mun&t=201900>

Idescat. (07 / 05 / 2020a). *Estructura per edats, envelliment i dependència*. Consultat el 01 / 03 / 2021, a Idescat: Indicadors demogràfics i de territori: <https://www.idescat.cat/pub/?id=inddt&n=915&geo=com:02#Plegable=geo>

Idescat. (28 / 08 / 2020b). *Mortalitat. Taxes per 100.000 habitants. Per tipus de malaltia, sexe i grups d'edat*. Recollit de Institut d'Estadística de Catalunya: <https://www.idescat.cat/pub/?id=aec&n=826>

Idescat. (05 / 02 / 2021a). *Població a 1 de gener. Per sexe i edat any a any*. Recollit de Institut d'Estadística de Catalunya: <https://www.idescat.cat/pub/?id=pmh&n=1180&geo=com:02>

Idescat. (15 / 01 / 2021b). *Taxa bruta de natalitat. Comarques i Aran, àmbits i províncies*. Consultat el 05 / 03 / 2021, a Idescat: <https://www.idescat.cat/pub/?id=aec&n=264>

Idescat. (15 / 01 / 2021c). *Naixements. Per edats de la mare. Comarques i Aran, àmbits i províncies*. Consultat el 05 / 03 / 2021, a Idescat: <https://www.idescat.cat/pub/?id=aec&n=260>

Indika Pol de Salut i Social de l'Alt Empordà. (01 / 01 / 2020). *Indika Pol de Salut i Social de l'Alt Empordà*. Recollit de Indika Pol de Salut i Social de l'Alt Empordà: <http://www.indikasalut.cat/>

Indika Pol de Salut i Social de l'Alt Empordà. (22 de 03 de 2021). *La Natalitat a l'Alt empordà*. Recuperado el 22 de 03 de 2021, de Indika:

Referències

- https://www.indikasalut.cat/wp-content/uploads/2021/03/INFOGRAFIA_natalitat_marc21.pdf
- Inkscape Project. (02 / 11 / 2003). Consultat el 22 / 03 / 2021, a Inkscape: <https://inkscape.org>
- Institute for Statistics and Mathematics of WU (Wirtschaftsuniversität Wien). (22 / 03 / 2021). Recollit de The comprehensive R Archive Network: <https://cran.r-project.org/>
- Jacobson, B., & Castillo-Salgado, C. (27 / 12 / 2014). *Providing health intelligence to meet local needs: a practical guide to serving local and urban communities through public health observatories*. Recollit de https://apps.who.int/iris/bitstream/handle/10665/152645/9789241508162_eng.pdf;sequence=1
- Kalra, D., Beale, T., & Heard, S. (2005). The openEHR foundation. *Studies in health technology and informatics*, 115, 153--173.
- Kinding, D., Bridget C., B., & Remington, P. (15 / 06 / 2010). Mobilizing Action Toward Community Health (MATCH): metrics, incentives, and partnerships for population health. *Preventing chronic disease*, 7(4), A68.
- Kneale, D., Khatwa, M., & Thomas, J. (2016). *Identifying and appraising promising sources of UK clinical, health and social care data for use by NICE*. (U. C. London, Ed.) London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education. Recollit de https://www.researchgate.net/profile/Dylan_Kneale/publication/308661443_Identifying_and_appraising_promising_sources_of_UK_clinical_health_and_social_care_data_for_use_by_NICE/links/57ea6ad408ae91a0c8d3c721/Identifying-and-appraising-promising-sources-of-
- Llano, J., García, J., Diego, J., & Recalde, J. (2014). *Aportación de los "Real World Data (RWD)" a la mejora de la práctica clínica*. Madrid, Espanya: Fundación Gaspar Casal. Recollit de <https://www.researchgate.net/profile/Juan-Llano->

Referències

- 2/publication/303881739_Aportacion_de_los_Real_World_Data_RWD_a_la_mejora_de_la_practica_clinica/links/575a822a08aed884620d43a3/Aportacion-de-los-Real-World-Data-RWD-a-la-mejora-de-la-practica-clinica.pdf
- ONU. (2020). Marco de indicadores mundiales para los Objetivos de Desarrollo Sostenible y metas de la Agenda 2030 para el Desarrollo Sostenible.
- Open Knowledge Foundation. (17 de 02 de 2021a). *Open Data Commons*. Recuperado el 17 de 02 de 2021, de Open Knowledge Foundation: <https://opendatacommons.org/licenses/pddl/1-0/>
- Open Knowledge Foundation. (17 de 02 de 2021b). *Open Data Handbook: What Is Open Data?* Recuperado el 17 de 02 de 2021, de Open Data Handbook: <http://opendatahandbook.org/guide/es/what-is-open-data/>
- R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. Vienna, Austria. Recollit de <https://www.R-project.org/>
- Reeder, B., & David, A. (2016). Health at hand: A systematic review of smart watch uses for health and wellness. *Journal of biomedical informatics*, 63, 269-276. doi:<https://doi.org/10.1016/j.jbi.2016.09.001>
- RStudio. (23 / 03 / 2021). *shinyapps.io*. Consultat el 23 / 03 / 2021, a shinyapps.io: <https://www.shinyapps.io/>
- Scarpato, N., Alessandra, P., Di Nunzio, L., & Fallucchi, F. (12 / 2017). E-health-IoT universe: A review. *International Journal on Advanced Science, Engineering and Information Technology*, 7. doi:10.18517/ijaseit.7.6.4467
- Simsion C, G., & Graham C, W. (2004). *Data modeling essentials* (Third Edition ed.). San Francisco: Elsevier.
- Staa, T.-P. v., Goldacre, B., Gulliford, M., Cassell, J., Pirmohamed, M., Taweel, A., . . . Smeeth, L. (2012). Pragmatic randomised trials using routine electronic health records: putting them to the test. *BMJ Publishing Group Ltd*, 344. doi:10.1136/bmj.e55

Referències

- Stachura, F., & Igras, K. (20 de 11 de 2020). *semantic.dashboard: Dashboard with Fomantic UI Support for Shiny*. Obtenido de semantic.dashboard: Dashboard with Fomantic UI Support for Shiny: <https://CRAN.R-project.org/package=semantic.dashboard>
- Stenberg, D. (20 de 11 de 2020). *CURL*. Obtenido de CURL: <https://curl.se/>
- The Association of British Pharmaceutical Industry. (2011a). *The vision for real world data. Harnessing the opportunities in the UK. Demonstrating value with real world data*. Londres, Regne Unit: ABPI.
- The Association of British Pharmaceutical Industry. (2011b). *Demonstrating value with real world data: A practical guide*. Londres, Regne Unit: ABPI.
- Whitehead M, D. G. (2006). *Levelling up (part 1): a discussion paper on concepts and principles for tackling social inequities in health*. (W. R. Europe, Ed.) Recuperado el 25 de 01 de 2021, de https://www.who.int/https://www.who.int/social_determinants/resources/leveling_up_part1.pdf
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. (S.-V. N. York, Ed.) Obtenido de <https://ggplot2.tidyverse.org>
- Wickham, H. (2019). *Advanced R*. Boca Raton: CRC Press. doi:<https://doi.org/10.1201/9781351201315>
- Wickham, H., Averick, M., Bryan, J., Chang, W., D'Agostino McGowan, L., François, R., . . . Milt, S. (2019). Welcome to the {tidyverse}. *Journal of Open Source Software*, 4(43), 1686. doi:10.21105/joss.01686
- Wilkinson, Richard G, Marmot, Michael, World Health Organization. Regional Office for Europe, WHO Centre for Urban Health (Europe) & International Centre for Health and Society. (2003). *Social determinants of health: the solid facts*. (R. W. Marmot, Ed.) Copenhagen: WHO Regional Office for Europe. Recollit de <https://apps.who.int/iris/handle/10665/108082>
- World Health Organization. (2009). *Subsanar Las Desigualdades En Una Generación: Alcanzar La Equidad Sanitaria Actuando Sobre Los*

Referències

- Determinantes Sociales De La Salud*. World Health Organization. Recollit de http://whqlibdoc.who.int/publications/2009/9789243563701_spa.pdf?ua=1
- World Health Organization. (2010). *A conceptual framework for action on the social determinants of health*. World Health Organization. Recollit de <https://apps.who.int/iris/handle/10665/44489>
- World Health Organization. (2016). *Commission on Social Determinants of Health*. Recuperado el 27 de 01 de 2021, de World Health Organization: http://www.who.int/social_determinants/thecommission/en/
- World Health Organization. (10 / 11 / 2020). *Social determinants of health*. Recollit de Social determinants of health: https://www.who.int/health-topics/social-determinants-of-health#tab=tab_1
- World Health Organization. (s.f.). *Enfermedades crónicas*. Recuperado el 09 de 12 de 2020, de World Health Organization: https://www.who.int/topics/chronic_diseases/es/
- World Health Organization. (s.f.). *Enfermedades no transmisibles*. Recuperado el 09 de 12 de 2020, de World Health Organization: https://www.who.int/topics/noncommunicable_diseases/es/
- World Health Organization. (s.f.). *Enfermedades no transmisibles y salud mental*. Recuperado el 09 de 12 de 2020, de World Health Organization: <https://www.who.int/nmh/topics/es/>

ANNEXOS

CODI DE LA INFOGRAFIA "POBLACIÓ A L'ALT EMPORDÀ"

```

#
# Codi de la infografia "Població a l'Alt Empordà"
#
# Carregar llibreries
library(tidyverse)
# Descarregar dades "Estructura per edats, envelliment i
dependència"
pag <- xml2::read_html(x =
"https://www.idescat.cat/pub/?geo=com%3A02iid=inddtin=915#Plegable=g
eo")
tab <- rvest::html_table(pag)
i <- tab[[1]][, 1] == 2019
x <- tab[[1]][i, ]

# Obtenir els indicadors d'interés.
PoblacioPediatria <- as.numeric(sub(",", ".", x[1,2], fixed =
TRUE))
PoblacioPediatria <- data.frame(
  category = c('P', 'NP'),
  count = c(PoblacioPediatria, 100 - PoblacioPediatria)
)

# Gràfic de donut (codi modificat de: https://www.r-graph-
gallery.com/128-ring-or-donut-plot.html)
ggplot(PoblacioPediatria, aes(x = 1, y = Count, fill = c('1',
'2')))) +
  geom_bar(stat = 'identity') +
  coord_polar('y')

# Compute percentages
PoblacioPediatria$fraction = PoblacioPediatria$count /
sum(PoblacioPediatria$count)

# Compute the cumulative percentages (top of each rectangle)
PoblacioPediatria$ymax = cumsum(PoblacioPediatria$fraction)

# Compute the bottom of each rectangle
PoblacioPediatria$ymin = c(0, head(PoblacioPediatria$ymax, n=-1))

# Make the plot

```

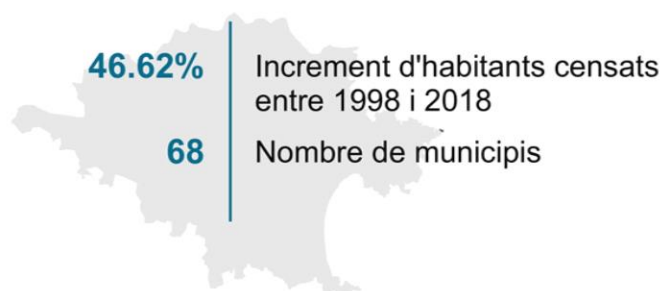

Annexos

```
ggplot(PoblacioPediatria, aes(ymax=ymax, ymin=ymin, xmax=4, xmin=3,
fill=category)) +
  geom_rect() +
  coord_polar(theta="y") + # Try to remove that to understand how
the chart is built initially
  xlim(c(2, 4)) + # Try to remove that to see how to make a pie
chart
  theme_void()
```

INFOGRAFIA "POBLACIÓ A L'ALT EMPORDÀ"

POBLACIÓ A L'ALT EMPORDÀ

149.569 habitants censats a l'Alt Empordà



El 16,27% de la població té menys de 15 anys.



El 18,20% de la població té més de 64 anys.



El 2,93% de la població té més de 84 anys.

Altres dades d'interés

El **61,85%** de la població de la comarca de l'Alt Empordà està censada a Figueres (33%), Roses (14%), Castelló d'Empúries (8%) i l'Escala (7%).

La **taxa bruta de natalitat per 1.000 habitants** a l'Alt Empordà l'any 2017 va ser de **9,80** un punt més alt que a **Catalunya** on al mateix període la taxa bruta de natalitat per 1.000 habitants va ser de **8,88** nens/habs.

A **Catalunya** la població amb una **edat inferior als 15** anys és del **15,49%**. La proporció de població amb més de 64 anys és de **18,70%** i la proporció de persones amb més de 84 anys és de **3,10%**

Font de les dades: Institut d'Estadística de Catalunya. amb les dades relatives a l'any 2018

CODI DE LA INFOGRAFIA “NATALITAT A L’ALT EMPORDÀ”

```

library(tidyverse)

df <- read.csv(
  '../01-data/02-tidy/01-parts-2020.csv',
  sep = ';',
  dec = ',',
  encoding = 'UTF-8'
)

Count <- sum(df$Num)

df <- df[1:10, ]

df$Pais <- factor(df$Pais, levels = rev(df$Pais))

ggplot(df[10:1, ], aes(x = Pais, y = 100 * Num / Count)) +
  geom_bar(aes(y = 100), fill = 'white', stat = 'identity') +
  geom_bar(stat = 'identity') +
  ylim(0, 100) +
  coord_flip()

ggsave(filename = '../04-svg/01-parts.svg')
df <- read.csv(file = '../01-data/01-raw/Consulta Edad Madres Pere
Plaja.csv', sep = ';', encoding = 'UTF-8')

df <- df %>% mutate(data = DNaixement, edat = Edat.mare) %>%
  select(data, edat)

df <- df %>% filter(edat != 227) %>% filter(edat > 3)

ggplot(df, aes(x = edat)) + geom_density() + theme_bw() +
  geom_vline(xintercept = 30) +
  geom_vline(xintercept = c(26, 34))
f <- list.files(path = '../01-data/01-raw/01-partos-por-mes-
idescat/', full.names = TRUE)

x <- lapply(f, read.csv, sep = ';', encoding = 'UTF-8', skip = 5)

names(x) <- 1975:2019

foo = function(y) {
  y <- y[, c(1, 4)]
}

```

Annexos

```
names(y) <- c("Mes", "N")
y %>% filter(Mes != 'Total')

}

z <- purrr::map_df(x, .f = foo, .id = 'Any')
z <- filter(z, Mes != 'No hi consta')

z$Mes <- factor(z$Mes, levels = c("Gener", "Febrer", "Març",
  "Abril", "Maig", "Juny", "Juliol", "Agost", "Setembre", "Octubre",
  "Novembre", "Desembre"))
z$Aux <- 1

ND <- data.frame(
  dia = seq.Date(
    from = as.Date('1975-01-01'),
    to = as.Date('2019-12-31'),
    by = 1
  )
)

ND <- ND %>%
  mutate(year = lubridate::year(dia), month = lubridate::month(dia))
%>%
  group_by(year, month) %>%
  summarise(ND = n(), .groups = 'drop') %>%
  select(ND)

z <- z %>%
  bind_cols(ND) %>%
  group_by(Any) %>%
  # mutate(Mean = mean(N)) %>%
  mutate(NE = sum(N) * ND / sum(ND)) %>%
  mutate(SD = sum(N) * (ND / sum(ND)) * ((sum(ND) - ND)/sum(N))) %>%
  mutate(PR = pnorm(N - NE, 0, 10.84317)) %>%
  mutate(CO = ifelse(PR > .5, '#202d80', '#ffffff'))

ggplot(z, aes(x = Any, y = Mes)) +
  geom_tile(color = '#918cb7', fill = z$CO)

# ggplot(z, aes(x = Any, y = Mes, fill = PR)) +
#   geom_tile(color = '#918cb7', fill = z$CO) +
#   scale_fill_gradientn(colours = c("#faf7ff", "#faf7ff",
# "#918cb7", "#202d80"))

mutate(Color = ifelse(N < NE, '#b5b5b5', '#202d80')) %>%
```

Annexos

```
mutate(Color = factor(Color, levels = c('#202d80', '#b5b5b5'))))

ggplot(z, aes(x = Any, y = 1, group = Mes)) +
  geom_bar(stat = 'identity', colour = '#202d80', fill = z$Color) +
  coord_polar('y')

ggplot(z, aes(x = Mes, y = Any)) +
  geom_tile(color = 'white', fill = z$Color) +
  theme_bw() +
  scale_x_discrete(expand = c(0,0)) +
  scale_y_discrete(expand = c(0,0)) +
  coord_polar()

# w <- z %>% group_by(Mes, Color) %>% summarise(N = n())
#
# ggplot(w, aes(x = Mes, y = N, fill = Color)) + geom_tile(stat =
#'identity') + coord_polar()
f <- list.files(path = '../01-data/01-raw/01-partos-por-mes-
idescat/', full.names = TRUE)

x <- lapply(f, read.csv, sep = ';', encoding = 'UTF-8', skip = 5)

names(x) <- 1975:2019

foo = function(y) {

  y <- y[, c(1, 4)]
  names(y) <- c("Mes", "N")
  y %>% filter(Mes == 'Total')

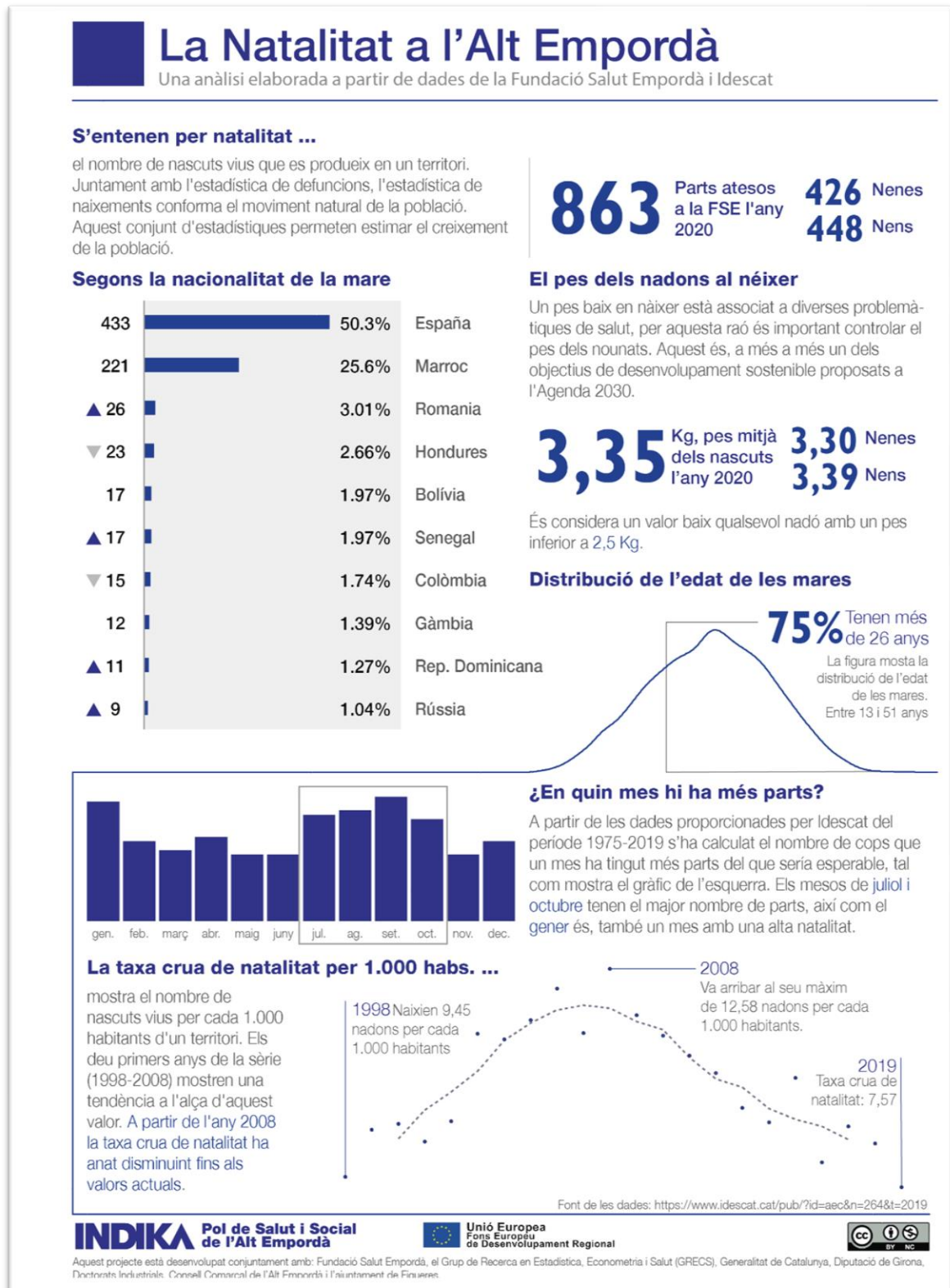
}

z <- purrr::map_df(x, .f = foo, .id = 'Any')
z <- z %>% filter(Any >= 1998)
z$POP <- c(
34546,
35028,
35443,
35777,
36639,
37744,
38824,
40017,
41081,
42586,
44178,
```

Annexos

```
45001,  
45326,  
45189,  
45299,  
44771,  
44578,  
44306,  
44107,  
43995,  
44065,  
44296)  
  
z <- mutate(z, TAXA = 1000 * N / POP)  
z <- mutate(z, MOVINGAVG = forecast::ma(TAXA, order = 5))  
  
z <- data.frame(  
  
  TAXA = c(  
    8.85, 9.74,  
    9.85, 9.51, 9.90, 11.55, 11.45,  
    11.81, 12.41, 11.57, 12.78, 11.90, 11.52, 11.14, 10.81,  
    10.15, 9.88, 10.72, 9.12, 9.80, 9.48, 8.65),  
  Any = c(1998:1999, 2000:2004, 2005:2009, 2010:2014, 2015:2019)  
)  
  
z <- mutate(z, MOVINGAVG = forecast::ma(TAXA, order = 5))  
  
ggplot(z, aes(x = as.integer(Any), y = TAXA)) +  
  geom_line(aes(y = MOVINGAVG), linetype = 'dashed', color =  
  '#787586', size = 1) +  
  geom_point(aes(y = TAXA), shape = 21, color = '#202d80', fill =  
  '#202d80', size = 3, ) +  
  theme_void()
```

INFOGRAFIA “NATALITAT A L’ALT EMPORDÀ”



CODI DE LA INFOGRAFIA “UN ANY DE COVID-19 A L’ALT EMPORDÀ”

```

# Codi infografia "Un any de COVID-19 a l'Alt Empordà"
## Load packages
library(tidyverse)
library(forecast)
library(RSocrata)

## Download municipality information
URL <- "https://analisi.transparenciacatalunya.cat/resource/jj6z-iyrp.json"
### Read credentials
credenciales <- readRDS(file = '01-data/credenciales.rds')

municipalities <-
  read.socrata(URL, app_token = credenciales$app_token, email =
credenciales$email, password = credenciales$password) %>%
  transmute(
    count = as.integer(numcasos),
    sex = as.factor(sexedescripcio),
    result = as.factor(resultatcoviddescripcio),
    data = as.Date(as.character(data)),
    region = comarcacodi
  ) %>% filter(result == "Positiu PCR")

## Catalonia
### Tidy data for Catalonia
catalonia <-
  municipalities %>%
  group_by(data) %>%
  summarise(count = sum(count)) %>%
  filter(data >= '2020-03-09') %>%
  filter(data <= '2021-03-09') %>%
  mutate(freq = 100000 * count / 7727029)

### Plot Moving Average
ggplot(catalonia, aes(x = data, y = ma(freq, 7) %>% as.numeric())) +
  geom_line() +
  theme_void()

## Alt Emporda
### Tidy data
alt <- municipalities %>%
  filter(region == "02") %>%
  group_by(data) %>%
  summarise(count = sum(count)) %>%

```


Annexos

```
filter(data >= '2020-03-09') %>%
filter(data <= '2021-03-09') %>%
mutate(freq = 100000 * count / 142624)

### Plot Moving Average with important dates has vertical lines.
ggplot(alt, aes(x = data, y = ma(freq, 7) %>% as.numeric())) +
  geom_line(colour = '#202d80') +
  geom_point(aes(y = freq), size = .75) +
  theme_void() +
  geom_vline(xintercept = as.Date('2020-03-14')) + # estat d'alarma.
  geom_vline(xintercept = as.Date('2021-02-14')) + # eleccions de
Catalunya.
  geom_vline(xintercept = as.Date('2020-06-21')) +
  geom_vline(xintercept = as.Date('2020-10-25')) +
  geom_vline(xintercept = as.Date('2020-12-24'), color = 'red') +
  geom_vline(xintercept = as.Date('2021-01-07'), color = 'red') +
  geom_vline(xintercept = as.Date('2021-01-14'), color = 'blue')

## Admissions
### Read Data
admissions <- read.csv(file = '01-data/01-raw/02-ingresos-per-
setmana.csv', sep = ';')
admissions_count <- sum(admissions$N)
### Plot admissions
ggplot(admissions, aes(x = 1:nrow(admissions), y = N)) +
  geom_line() +
  theme_void()

## Deaths
### Download Data
URL <- "https://analisi.transparenciacatalunya.cat/resource/uqk7-
bf9s.json"

deaths <-
  read.socrata(URL,
               app_token = credenciales$app_token,
               email = credenciales$email,
               password = credenciales$password
  ) %>%
  filter(comarcacodi == '02') %>%
  group_by(exitusdata) %>%
  summarise(N = as.integer(numexitus)) %>%
  ungroup() %>%
  filter(exitusdata >= as.Date('2020-03-09')) %>%
  filter(exitusdata <= as.Date('2021-03-09'))
```

Annexos

```
aux <- data.frame(
  exitusdata = seq.Date(
    as.Date('2020-03-01'),
    as.Date('2021-03-01'),
    by = 1
  )
)

deaths <- left_join(aux, deaths) %>%
  replace_na(replace = list(N = 0))

deaths_count <- sum(deaths$N)

monthly_count <- deaths %>%
  mutate(
    year = lubridate::year(exitusdata),
    month = lubridate::month(exitusdata)
  ) %>%
  group_by(year, month) %>%
  summarise(N = sum(N))

ggplot(monthly_count, aes(x = 1:13, y = N)) +
  geom_bar(stat = 'identity', fill = '#202d80') +
  theme_void()

## Infections by age
URL <- "https://analisi.transparenciacatalunya.cat/resource/qwj8-
xpvk.json"

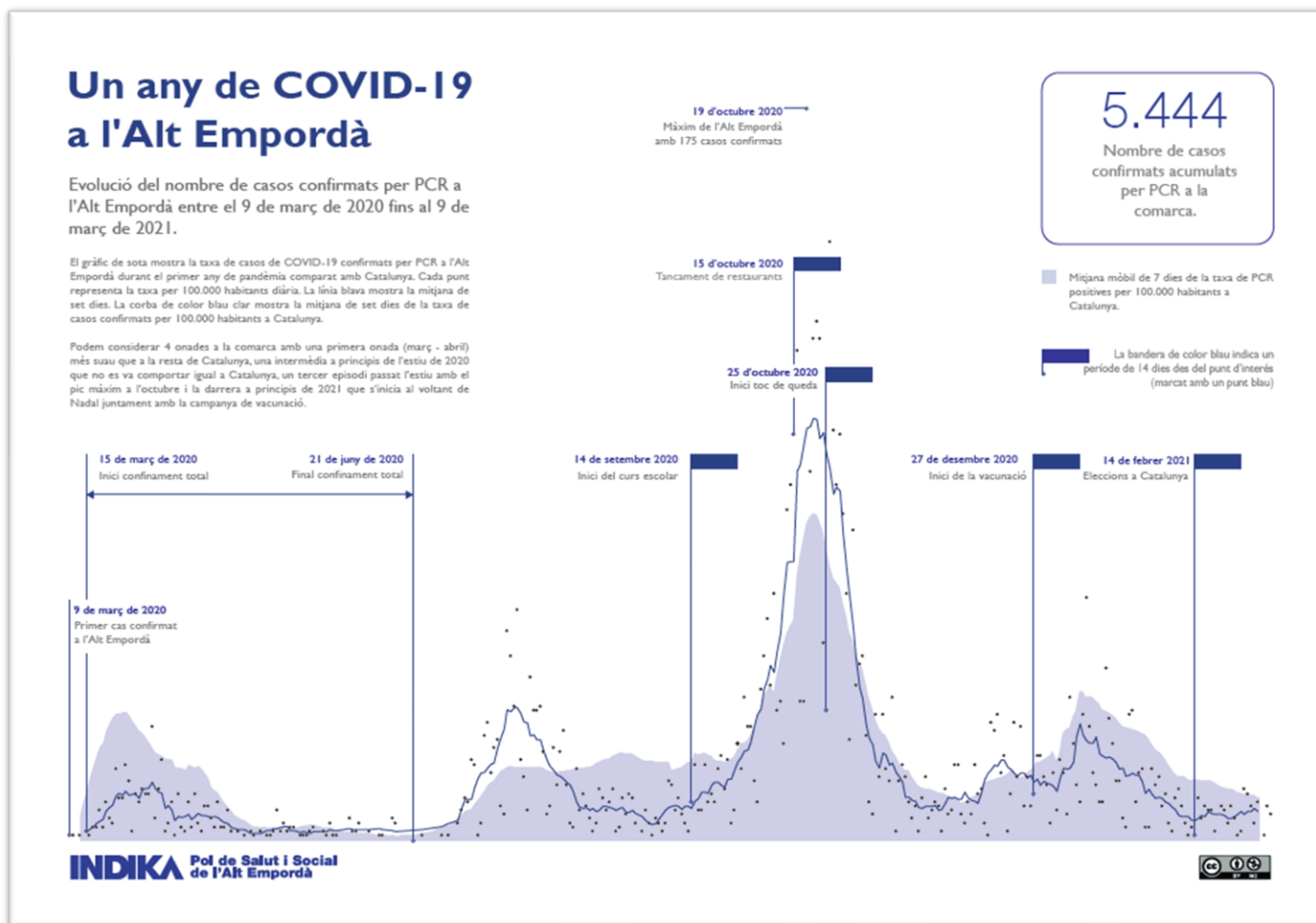
infections <-
  read.socrata(URL,
    app_token = credenciales$app_token,
    email = credenciales$email,
    password = credenciales$password
  ) %>%
  filter(regiosanitariacodi == '6400') %>%
  group_by(edatrang) %>%
  summarise(N = sum(as.integer(numcasos)))

### Populations by age group
pop <- data.frame(
  edatrang = c("0-9", "10-19", "20-29", "30-39", "40-49", "50-59",
    "60-69", "70-79", "80-89", "90+"),
  popu = c(78806, 87126, 84624, 103319, 130224, 110865, 85568,
    59791, 32780, 8685)
```

Annexos

```
)  
  
### Compute 100000 freq  
infections <- left_join(infections, pop) %>%  
  mutate(taxa = 100000 * N / popu)  
  
ggplot(infections[1:10, ], aes(x = edatrang, y = taxa)) +  
  # geom_bar(aes( y = 100000), fill = 'grey', stat = 'identity') +  
  geom_bar(stat = 'identity', fill = '#202d80') +  
  # ylim(0, 100000) +  
  theme_void()
```

INFOGRAFIA “UN ANY DE COVID-19 A L’ALT EMPORDÀ”



Nombre d'ingressos hospitalaris totals

La corva d'ingressos ha estat paral·lela a la de casos comarcals (amb petit impasse) ha presentat 4 pics (amb petit desescalatge). Març, juny i octubre de 2020 i Gener de 2021.

Font: Dades de la Fundació Salut Empordà

1.098

Nombre total d'ingressos a la FSE per COVID-19

Nombre de defuncions

El nombre de defuncions s'ha incrementat en el segon semestre analitzat, coincidint amb el 3er i 4t brot.

Font: <https://analisi.transparenciacatalunya.cat/resource/uqk7-bf9s.json>

224

Nombre total de defuncions a l'Alt Empordà.

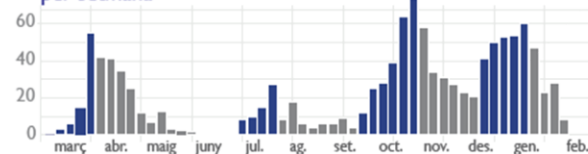
Nombre de casos per edat a Girona

La corva de les edats de la covid esta condicionada per una major detecció en persones grans en el primer període i una disminució de en les edats del segon . probablement afavorit per els screenings associats a àmbit escolar i la millora en les residències.

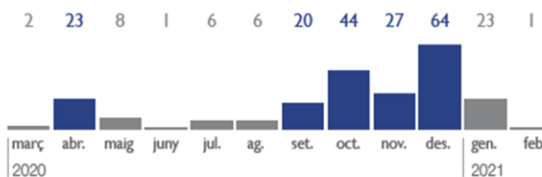
Edat	Núm. casos	%
0-9	3.827	6,24
10-19	7.759	12,60
20-29	7.141	11,60
30-39	8.157	13,30
40-49	10.508	17,10
50-59	8.780	14,30
60-69	5.771	9,41
70-79	3.836	6,25
80-89	3.534	5,76
90+	1.870	3,05

Font: <https://analisi.transparenciacatalunya.cat/resource/qwj8-xpvk.json>

Nombre de nous ingressos hospitalaris per COVID-19 per setmana



Nombre de defuncions mensuals



Estat de la vacunació

A partir del 27 de desembre de 2020 es va iniciar el procés de vacunació a l'Alt Empordà. Al finalitzar el període d'aquesta infografia el nombre de vacunats a l'Alt Empordà és de:

2.411

Nombre total de vacunats amb la segona dosi a l'Alt Empordà

1.69%

Percentatge de població vacunada a l'Alt Empordà

Font: <https://analisi.transparenciacatalunya.cat/resource/cuwj-bh3b.json>



CODI DE LA INFOGRAFIA “UN ANY DE COVID-19 A L’ALT EMPORDÀ”

```

# Codi infografia "Un any de COVID-19 a l'Alt Empordà"
## Load packages
library(tidyverse)
library(forecast)
library(RSocrata)
library(ggmosaic)

## Download municipality information
URL <- "https://analisi.transparenciacatalunya.cat/resource/jj6z-iyrp.json"

### Read credentials
credenciales <- readRDS(file = '01-data/credenciales.rds')

municipalities <-
  read.socrata(URL, app_token = credenciales$app_token, email =
credenciales$email, password = credenciales$password) %>%
  transmute(
    count = as.integer(numcasos),
    sex = as.factor(sexedescripcio),
    result = as.factor(resultatcoviddescripcio),
    data = as.Date(as.character(data)),
    region = comarcacodi
  ) %>% filter(result == "Positiu PCR")

## Catalonia
### Tidy data for Catalonia
catalonia <-
  municipalities %>%
  group_by(data) %>%
  summarise(count = sum(count)) %>%
  filter(data >= '2020-03-09') %>%
  filter(data <= '2021-03-09') %>%
  mutate(freq = 100000 * count / 7727029)

### Plot Moving Average
ggplot(catalonia, aes(x = data, y = ma(freq, 7) %>% as.numeric())) +
  geom_line() +
  geom_vline(xintercept = as.Date("2020-03-09")) +
  geom_vline(xintercept = as.Date("2021-03-09")) +
  geom_hline(yintercept = 122.7002) +
  geom_hline(yintercept = 0) +
  theme_void()

```

Annexos

```
## Alt Emporda
### Tidy data
alt <- municipalities %>%
  filter(region == "02") %>%
  group_by(data) %>%
  summarise(count = sum(count)) %>%
  filter(data >= '2020-03-09') %>%
  filter(data <= '2021-03-09') %>%
  mutate(freq = 100000 * count / 142624)

### Plot Moving Average with important dates has vertical lines.
ggplot(alt, aes(x = data, y = ma(freq, 7) %>% as.numeric())) +
  geom_line(colour = '#202d80') +
  geom_point(aes(y = freq), size = 1, shape = 21, fill = 'black') +
  theme_void() +
  geom_vline(xintercept = as.Date('2020-03-14')) + # estat d'alarma.
  geom_vline(xintercept = as.Date('2020-06-21')) +
  geom_vline(xintercept = as.Date('2020-10-15')) +
  geom_vline(xintercept = as.Date('2020-10-25')) + # inici toc de
queda
  geom_vline(xintercept = as.Date('2020-12-24')) +
  geom_vline(xintercept = as.Date('2020-12-27')) + # inici vacunació
  geom_vline(xintercept = as.Date('2021-01-07')) +
  geom_vline(xintercept = as.Date('2021-01-14')) +
  geom_vline(xintercept = as.Date('2021-02-14')) + # eleccions de
Catalunya.
  geom_vline(xintercept = as.Date("2020-03-09"), color = 'red') +
  geom_vline(xintercept = as.Date("2021-03-09"), color = 'red') +
  geom_vline(xintercept = as.Date("2020-09-14")) +
  geom_vline(xintercept = as.Date("2020-09-28")) +
  geom_hline(yintercept = 0)

alt_count <- alt$count %>% sum()

## Admissions
### Read Data
admissions <-
  read.csv(file = '01-data/01-raw/02-ingresos-per-setmana.csv', sep
= ';') %>%
  mutate(week = setmana, year = any) %>%
  select(year, week, NADMIN = N)

admissions_count <- sum(admissions$N)

### Plot admissions
```

Annexos

```
ggplot(admissions, aes(x = 1:nrow(admissions), y = NADMIN)) +
  geom_bar(stat = 'identity') +
  theme_minimal() +
  xlab(element_blank()) +
  ylab(element_blank())

## Deaths
### Download Data
URL <- "https://analisi.transparenciacatalunya.cat/resource/uqk7-
bf9s.json"

deaths <-
  read.socrata(URL, app_token = credenciales$app_token, email =
credenciales$email, password = credenciales$password) %>%
  filter(comarcacodi == '02') %>%
  group_by(exitusdata) %>%
  summarise(N = as.integer(numexitus)) %>%
  ungroup() %>%
  filter(exitusdata >= as.Date('2020-03-09')) %>%
  filter(exitusdata <= as.Date('2021-03-09'))

aux <- data.frame(
  exitusdata = seq.Date(
    as.Date('2020-03-01'),
    as.Date('2021-03-01'),
    by = 1
  )
)

deaths <- left_join(aux, deaths) %>%
  replace_na(replace = list(N = 0))

deaths_count <- sum(deaths$N)

monthly_count <- deaths %>%
  mutate(
    year = lubridate::year(exitusdata),
    month = lubridate::month(exitusdata)
  ) %>%
  group_by(year, month) %>%
  summarise(N = sum(N))

ggplot(monthly_count, aes(x = 1:12, y = N)) +
  geom_bar(stat = 'identity', fill = '#202d80') +
  theme_void()
```


Annexos

```
## Infections by age
URL <- "https://analisi.transparenciacatalunya.cat/resource/qwj8-
xpvk.json"

infections <-
  read.socrata(URL,
               app_token = credenciales$app_token,
               email = credenciales$email,
               password = credenciales$password
  ) %>%
  filter(data >= '2020-03-09') %>%
  filter(data <= '2021-03-09') %>%
  filter(regiosanitariacodi == '6400') %>%
  group_by(edatrang) %>%
  summarise(N = sum(as.integer(numcasos))) %>%
  mutate(NF = 100 * N / sum(N)) %>%
  .[-11, ]

ggplot(infections, aes(x = edatrang, y = N)) +
  geom_bar(fill = 'grey', aes(y = rep(100,10)), stat = 'identity') +
  geom_bar(fill = "#202d80", stat = 'identity') +
  coord_flip() +
  theme_void()

# Vaccination
URL <- "https://analisi.transparenciacatalunya.cat/resource/cuwj-
bh3b.json"

vaccined <- read.socrata(URL, app_token = credenciales$app_token,
                        email = credenciales$email, password = credenciales$password)
vaccined <- vaccined %>%
  filter(comarca == 'ALT EMPORDA') %>%
  filter(data <= as.Date("2021-03-09")) %>%
  filter(!is.na(fabricant)) %>%
  group_by(dosi) %>%
  summarise(N = sum(as.integer(recompte)))
```

KEY FIGURES ON EUROPE

Aquesta taula mostra la revisió d'indicadors disponibles segons el seu nivell de desagregació territorial a partir del llistat d'indicadors presentat en el document *Key Figures On Europe* (EUROSTAT, 2019). El color blau representa que l'indicador corresponent està disponible per al nivell de desagregació indicat en la capçalera de la columna que pot ser: Europa (EU), Espanya (ES), Catalunya (CAT), Província de Girona (GIR), Comarca de l'Alt Empordà (ALT) i Municipis de l'Alt Empordà (MUN).

Capítol	Secció	Indicador	Disponibilitat					
			EU	ES	CAT	GIR	ALT	MUN
People and society	Population	Population	■	■	■	■	■	■
		Population developments	■	■	■	■	■	■
		Population change	■	■	■	■	■	■
		Fertility rate	■	■	■			

Annexos

		Ageing population						
		Population by citizenship						
		First time asylum application						
	Health	Developments for life expectancy at birth						
		Life expectancy at birth						
		Main causes of death						
		Deaths from diseases of the circulatory system						
	Education	Early leavers from education and training						

Annexos

		Learning two or more foreign languages	■	■	■			
		Young people neither in employment nor in education and training	■	■	■			
		Fields of study for university graduates	■	■	■			
	Labour market	Labour force composition	■					
		Employment rate	■	■	■	■	■	■
		Unemployment rate development	■	■	■	■	■	■
		Unemployment rate	■	■	■	■	■	■

Annexos

		Gender pay gap						
	Living condition	People at risk of poverty or social exclusion						
		People at risk of poverty or social excluion						
		Inability to face unexpected financial expenses						
		Inability to afford paying for one week anual holiday away from home						
	Digital society	Internet usage						
		Internet activities						

Economy and business	Economy and finance	GDP						
		Real change in GDP						
		GDP per capita						
	Prices	Inflation rate						
		Household budget structure						
		Comparative price levels						
	Government finance	General government deficit/surplus and debt						
		Structure of general government expenditure						

		International trade in goods with non-member countries					
		Top 10 partners for EU International trade in goods					
	International trade	International trade in goods with Member States					
		International trade in services with non-member countries					
		Top 10 partners for EU International trade in services					

Annexos

		Change in the structure of value added						
		Gross value added by sector						
	Business	Change in the structure of employment						
		Employment by sector						
		Non-financial business economy						
		Value added in the non-financial business economy						

Annexos

		Developments for industrial output						
	Research and development	Gross domestic expenditure on R i D						
	Tourism	Nights spent in EU tourist accommodation						
		Top 20 EU regions in terms of nights spent in tourist accommodation						
		Nights spent in tourist accommodation						
	Environment and natural resources	Transport	Passenger cars					
Age of passenger cars								

		Air passenger transport	■	■				
		Top 20 EU airport	■	■				
		Inland freight transport by type of transport	■	■				
		Top 20 EU short sea shipping ports	■	■				
	Energy	Origin of energy imports	■	■				
		Energy dependency rate	■	■				
		Energy from renewable energy source	■	■	■	■		
		Natural gas and electricity prices	■	■	■			

Annexos

		Structure of final energy consumption						
		Final energy consumption by end use						
	Environment	Employment in the environmental economy						
		Environmental tax revenue						
		Greenhouse gas emissions						
		Greenhouse gas emissions by source sector						

Annexos

		Municipal waste generation						
		Municipal waste treatment methods						
	Agriculture	Top 5 EU Member States with the largest production for selected agricultural products						
		Organic crop area						
	Fisheries	EU Member States with the largest fish catches						
	Forestry	Area of forest and other wooded land						

LLISTAT DELS INDICADORS CLAU DE L'ALT EMPORDÀ

		Disponibilitat				
		CAT	GIR	ALT	URL	
Persones i població	Població	Població			http://www.idescat.cat/pub/?id=pmhin=446igeo=com:02iby=mun	
		Evolució de la població				http://www.idescat.cat/pub/?id=ep
		Canvi en la població				http://www.idescat.cat/pub/?id=ep
		Taxa bruta de natalitat				http://www.idescat.cat/pub/?id=aecin=264

Annexos

		Envel·liment de la població				http://www.idescat.cat/pub/?id=proj
		Població: per sexe i quinquennal				http://www.idescat.cat/pub/?id=pmhin=9548igeo=com:02ilang=es
		Població per ciutadania				http://www.idescat.cat/pub/?id=pmhin=679iby=munigeo=com%3A02#Plegable=geo
Salut		Evolució de l'esperança de vida en néixer				http://www.idescat.cat/pub/?id=ievin=8619igeo=prov:17#Plegable=geo
		Principals causes de mortalitat				Dades de la Fundació Salut Empordà
Educació		Taxa d'escolarització als 17 anys				http://www.idescat.cat/pub/?geo=com%3A02iid=intpobrin=8233#Plegable=geo
Força de treball		Composició de la força de treball				http://www.idescat.cat/pub/?id=censphin=270iby=com

Annexos

		Evolució de l'atur registral				http://observatorideltreball.gencat.cat/ObservatoriDeITreball/servlet/mstrWeb
		Taxa d'atur registral				https://observatoritreball.gencat.cat/ca/inici
	Condicions de vida	Desigualtat de renda segons fonts tributàries				http://www.idescat.cat/pub/?id=intpobrin=8228iby=com
		Percentatge de rendes inferior al 60% de la mediana				http://www.idescat.cat/pub/?id=intpobrin=8228iby=com
		Bretxa de les rendes inferiors al 60% de la mediana				http://www.idescat.cat/pub/?id=intpobrin=8228iby=com
		Bretxa salarial de gènere				http://www.idescat.cat/indicadors/?id=anualsin=10405

	Societat Digital	Ús d'internet				http://www.idescat.cat/pub/?id=aecin=621
		Activitat a internet				http://www.idescat.cat/pub/?id=ticll19in=3.3.1
Economia i negocis	Economia i finances	Producte interior brut				http://www.idescat.cat/pub/?id=pibcin=8276iby=comilang=es
		Taxa de variació del PIB real				http://www.idescat.cat/pub/?geo=com%3A02iid=pibcin=8276#Plegable=geo
		PIB per càpita a l'Alt Empordà				http://www.idescat.cat/pub/?id=pmhin=1181igeo=com:02ilang=es http://
	Preus	Despesa de les llars				http://www.idescat.cat/pub/?id=edclin=9493
	Finances del govern	Dèficit/excedent i deute del govern				www.seu-e.cat/ca/web/ccaltemporda/govern-obert-i-transparencia/gestio-economica/gestio-economica/endeutament

Annexos

		Estructura de la despesa pública				www.seu-e.cat/ca/web/ccaltemporda/govern-obert-i-transparencia/gestio-economica/pressupost/liquidacio-del-pressupost/despeses-per-programa
	Negocis	Canvis en l'estructura del valor afegit brut				http://www.idescat.cat/pub/?geo=com:02iid=pibcin=8277#Plegable=geo
		Població ocupada per branca d'activitat				http://www.idescat.cat/pub/?id=aecin=312
	Turisme	Viatgers en establiments hotelers				Demandar a IDESCAT
		Pernoctacions en establiments hotelers				Demandar a IDESCAT
		Graus d'ocupació en establiments hotelers				Demandar a IDESCAT
Entorn i	Transport	Parc de vehicles per tipus				http://www.idescat.cat/pub/?id=aecin=588

Annexos

		Edat del parc de vehicles				Demanar a la DGT
	Energia	Estructura final del consum d'energia elèctrica				Demanar a l'Institut Català d'Energia
		Estructura final del consum de gas natural				Demanar a l'Institut Català d'Energia
	Entorn	Generació de residus municipals				http://www.idescat.cat/pub/?id=resmcin=6997igeo=com:02
		Tipus de residus				http://www.idescat.cat/pub/?id=resmcin=7000igeo=com:02 http://www.idescat.cat/pub/?id=resmcin=6998igeo=com:02
	Agricultura	Ús de la superfície per a conreu segons tipus				http://www.idescat.cat/pub/?id=aecin=202

Annexos

	Silvicultura	Usos del sòl				http://www.idescat.cat/pub/?id=aecin=202
--	--------------	--------------	--	--	--	---

Annexos