Proposal of a Learning Health System to Transform the National Health System of Spain

Authors:

Rafael Carnicero, David Rojas, Ignacio Elicegui, Javier Carnicero

Abstract:

This article identifies the main challenges of the National Health Service of Spain and proposes its transformation into a Learning Health System. For this purpose, the main indicators and reports published by the Spanish Ministries of Health and Finance, Organization for Economic Co-operation and Development (OECD) and World Health Organization (WHO) were reviewed. The Learning Health System proposal is based on some sections of an unpublished report, written by two of the authors under request of the Ministry of Health of Spain on Big Data for the National Health System. The main challenges identified are the rising old age dependency ratio; health expenditure pressures and the likely increase of out-of-pocket expenditure; drug expenditures, both retail and consumed in hospitals; waiting lists for surgery; potentially preventable hospital admissions; and the use of electronic health record (EHR) data to fulfil national health information and research objectives. To improve its efficacy, efficiency, and quality, the National Health Service of Spain should be transformed into a Learning Health System. Information and communication technologies (IT) enablers are a fundamental tool to address the complexity and vastness of health data as well as the urgency that clinical and management decisions require. Big Data solutions are a perfect match for that problem in health systems.

Proposal of a Learning Health System to Transform the National Health System of Spain

Rafael Carnicero [1], David Rojas [2], Ignacio Elicegui [3] and Javier Carnicero [4,*]

[1] Mateprincs SL, 31200 Estella, Spain
[2] SATEC, Aravaca, 28023 Madrid, Spain
[3] Communications Engineering Department, University of Cantabria, 39005 Santander, Spain
[4] Cognitio Salud, 31007 Pamplona, Spain
[*] Correspondence: javiercarnicero@cognitiosalud.es; Tel.: +34-697-902-000

Abstract: This article identifies the main challenges of the National Health Service of Spain and proposes its transformation into a Learning Health System. For this purpose, the main indicators and reports published by the Spanish Ministries of Health and Finance, Organization for Economic Co-operation and Development (OECD) and World Health Organization (WHO) were reviewed. The Learning Health System proposal is based on some sections of an unpublished report, written by two of the authors under request of the Ministry of Health of Spain on Big Data for the National Health System. The main challenges identified are the rising old age dependency ratio; health expenditure pressures and the likely increase of out-of-pocket expenditure; drug expenditures, both retail and consumed in hospitals; waiting lists for surgery; potentially preventable hospital admissions; and the use of electronic health record (EHR) data to fulfil national health information and research objectives. To improve its efficacy, efficiency, and quality, the National Health Service of Spain should be transformed into a Learning Health System. Information and communication technologies (IT) enablers are a fundamental tool to address the complexity and vastness of health data as well as the urgency that clinical and management decisions require. Big Data solutions are a perfect match for that problem in health systems.

Keywords: learning health system; Big Data applied to healthcare; Big Data applied to clinical research; data protection; patient's phenotyping; Big Data applied to National Health System of Spain

1. Introduction

1.1. Health Systems' Challenges

Health systems in western European countries currently deal with several serious challenges, such as population aging, increasing impact of chronic diseases, rising demand on healthcare services, the process of continuous innovation that brings expensive new healthcare technologies, and, as a result of all of these, an increase in public expenditure that must coexist with the corresponding budget constraints. Moreover, western healthcare systems present some shortcomings which cannot be addressed exclusively by advances in science: failure to address preventable diseases; low safety level of patients; unwanted variations in quality of care; lack of efficiency in the use of resources; and inequity in treatment access and results. These deficiencies contribute to the growth of public health expenditure and have a high social cost, which comes at the same time as a lower public outlay, in both gross as well as percentage points of the GDP [1].

Quality, efficiency, and effectiveness enhancement demands that health systems focus on the needs and preferences of individual patients, and not on patients as an abstract concept. Therefore,

there must be specific, measurable, and personally meaningful goals for each patient. This means they must have clinical and individual value for each person. In medical terms, value can be defined as a clinical status improvement, including the quality of life of each patient [2]. Medical communication and technical quality of care, accessibility and physical environment satisfaction, interpersonal manner, and empathy are some of the dimensions that should be measured [3].

1.2. Information and Communication Technologies Are the Tool to Transform Health System

The aforementioned challenges require health systems to increase their effectiveness and efficiency. This process must rely heavily on information and communication technologies (IT), since any healthcare activity needs an intensive processing of information. IT brings improvements in efficiency and quality of operations management in health systems by reducing operating costs of clinical services, reducing administrative costs, and transforming healthcare. IT implementation can result in care that is both higher in quality, safer, and more responsive to patients' needs and, at the same time, more efficient (appropriate, available, and less wasteful) [4].

Improving the effectiveness and efficiency of the process requires acquiring and applying new knowledge. Health systems generate knowledge and produce data which, if duly analyzed, will create new knowledge. This knowledge management includes the practice of data mining, which can be highly beneficial for improving health systems. Therefore, generating value in the health system should be based on taking advantage of existing and created (by the system) knowledge. As a rule of thumb, everything that does not generate value is a cost, and must be eliminated.

Health data constitutes a significant resource and it makes economic and ethical sense to use this data as much as possible to improve the health of the population and the effectiveness, safety, and patient-centeredness of healthcare systems. Rising levels of chronic disease and multi-morbidity, concerns about the quality and safety of patient care, the need to measure and assure value for money for investments in health, and the need to allocate health system resources require good evidence for decision making (Figure 1) [5].
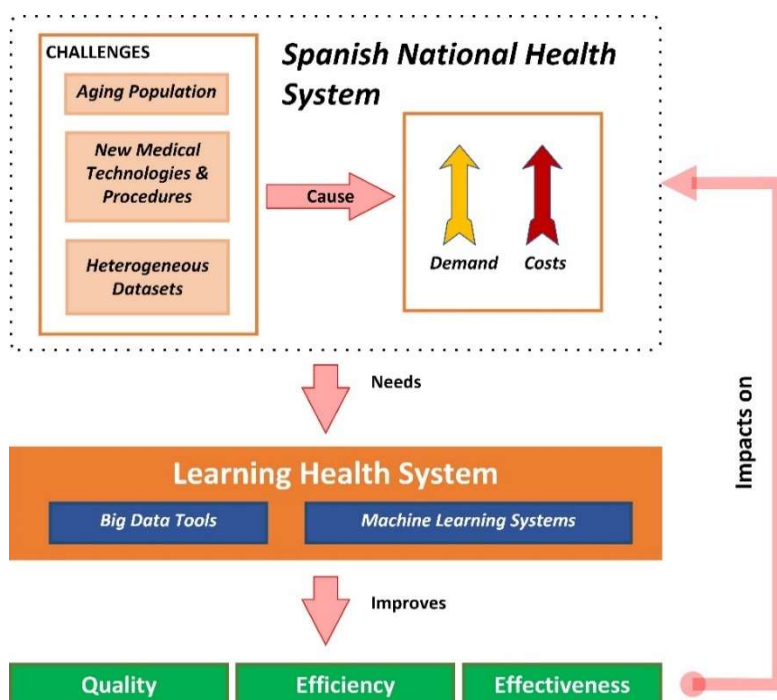

Figure 1. Challenges of health systems and opportunities of learning health systems.

Health system managers and policy makers are moving towards performance-based governance to improve care quality, co-ordination, and efficiency. Performance-based governance requires timely

and accurate patient data that ensure continuity in care, including health outcomes and costs. Such data also support the re-design and evaluation of new models of healthcare service delivery and contribute to the discovery and evaluation of new treatments [6].

1.3. Objectives

This article has three purposes:

1. Identify the current challenges of the Spanish National Health System.
2. Propose a new Big Data based solution which will improve the quality, efficiency, and effectiveness of the Spanish National Health System to address these challenges. This process needs to use both clinical and non-clinical data to aid the decision-making processes related to patients' health, research, and evaluation of the performance of the health system. The governance, architecture, legal framework, bioethical principles, data homogeneity, and technological tools needed for its creation, operation, and evolution must also be considered.
3. Propose the creation of a Learning Health System (LHS) for the National Health System of Spain, envisioned as an integrated framework "in which progress in science, informatics, and care culture align to generate new knowledge as an ongoing, natural by-product of the care experience, and seamlessly refine and deliver best practices for continuous improvement in health and healthcare" [7].

2. Materials and Methods

To identify the Spanish National Health System current challenges and analyze its performance, statistics and reports from the Spanish Ministry of Health, the Spanish National Statistics Institute, the Spanish Ministry of Finance and the Organization for Economic Co-operation and Development (OECD) were studied, as well as the reports provided by the National Health System Observatory (Ministry of Health). The selected dataset analyzed relates to dependence, life expectancy, mortality, chronic diseases (preventable hospitalizations) health, and pharmaceutical expenditure. In addition, reports from the OECD and World Health Organization (WHO) showing efficiency, Big Data and IT usage in health systems have been considered. In all cases results obtained in Spain were compared with those reported by the European Union (EU) and the OECD member countries. The progress in national dataset availability and health data governance based on OECD reports was also reviewed.

The Learning Health System proposal is based on the sections of an unpublished Big Data for the National Health System report written by two of the authors under request of the Ministry of Health of Spain.

Moreover, non-systematic searches in the PubMed database using the key words "Big Data and Health" and "Big Data and Healthcare" have been carried out.

3. Results

In Spain, the National Health System provides universal care and is funded by the government via taxes. The management of the healthcare services, as part of the National Health System, is decentralized, relying on each Regional Government, although the Ministry of Health is responsible for the systems and information technologies shared by all the health services. An Inter-territorial Board of the National Health System oversees the coordination between the healthcare providers in which all the 17 regions take part. Such a board is tasked with acquiring knowledge, debating, and advising on the essential functions of the National Health System, and planning, coordinating, and evaluating its performance.

The National Health System of Spain is considered as one of the top performing ones in terms of outputs, such as life expectancy, as well as performance, defined as the best result that can be obtained with an equal amount of resources. The World Health Organization ranked it as the seventh best performing country among 191 countries in a report published in the year 2000 [8].

Table 1 summarizes the main indicators obtained from the review conducted on dependence, life expectancy, spending in healthcare, and share of avoidable admissions.

Table 1. Indicators on dependence, life expectancy, spending in healthcare and share of avoidable admissions due to five chronic diseases.

Indicator	Spain	European Union
Population under 15 years old/Total population	15%	16%
Population older than 65/Total population	19.0%	19.0%
Life expectancy at birth (years)	82.8	81.0
Life expectancy at birth (women) (years)	86.3	83.6
Healthy life years (women)	66.5	64.2
Life expectancy at birth (men) (years)	80.5	78.2
Healthy life years (men)	65.9	63.5
Spending on healthcare (% GDP)	8.8	9.6
Share of total health spending financed by out-of-pocket payments	23.8%	18.2%
Pharmaceutical expenditure (retail) per capita and as a share of health expenditure	19.1%	16.8%
Pharmaceutical expenditure (retail) per capita financed by government/compulsory schemes as a share of pharmaceutical expenditure	57%	64%
Share of potentially avoidable hospital admissions due to five chronic conditions: diabetes, hypertension, heart failure, COPD and bronchiectasis, and asthma.	6.3%	5.5%

Source: Organization for Economic Co-operation and Development (OECD) /European Union (EU) (2018), Health at a Glance: Europe 2018: State of Health in the EU Cycle. [9]

3.1. Dependency

The dependency ratio is the demographic index that shows as a percentage the relation between dependent population (defined as younger than 16 and older than 64) and productive population (between 16 and 64). This ratio can be divided between youth dependency rate (under 16) and senior dependency rate (over 64). The dependency rate in Spain is 53.4%, comprised of a senior dependency rate of 28.7% and a 24.7% young dependency rate [10]. In the EU, 16% of its population is under 15 years old and 19% is above 65. The country with the highest population under 15 years old is Ireland, while Germany, with 13%, has the lowest. Spain, with 15% of its population under 15 years old, and 19% older than 65, has an average rate (Table 1) [9].

In Spain, the population over 65 years old increased between 2003 and 2017 from 16.96% of the total to 18.96%, while the senior dependency rate increased from 24.82% to 29.17% during those same years. The ratio in 2019 is 29.94% (provisional) [11].

3.2. Life Expectancy at Birth and Healthy Life Expectancy

Life expectancy at birth measures the average number of years that a person can expect to live based on current mortality rates (age-specific death rates). Life expectancy at birth in Spain in 2016 was 82.8 years, 80.0 for men and 86.3 for women, which is the highest in the EU. Between 2001 and 2015, it increased by 3.1 years (3.7 for men and 2.5 for women). The leading causes of death continue to be diseases of the circulatory system (which represent 29.7% of total deaths) and tumors (26.8% of total deaths) [9,10].

Healthy life expectancy is an important indicator of population health. It indicates whether any gains in life expectancy are lived in good health or with some health problems and disabilities. A greater number of healthy life years generally means a healthier workforce, fewer early retirements due to health problems, and reduced long-term care needs. The main indicator of healthy life years used in the EU is the number of years lived free of activity limitations due to health problems (in other words, disability-free life expectancy). The healthy life expectancy in Spain is higher than the EU average (Table 1) [9].

3.3. Mortality and Morbidity

The main causes of death in EU member countries are diseases of the circulatory system (38%), malignant tumors (26%), and external causes (5%). Spain is among the EU28 countries with the lowest rate of death from ischemic heart disease, cerebral stroke or breast cancer. The Spanish rate of death from laryngeal cancer and windpipe and respiratory diseases is lower than the EU average. In addition, Southern European countries—Greece, Cyprus, Malta, Italy, and Spain—jointly with the United Kingdom are the regions showing the lowest rate of death from suicide, while Eastern European countries (Slovenia, Latvia, Hungary, and Lithuania) have the highest suicide rates [9].

3.4. Health Expenditure

Health expenditure in Spain in 2005 was 7.7% of the total GDP and in 2017 it represented 8.8%. Nevertheless, during 2011, 2012, and 2015, this value increased to 9.1% of the GDP, as shown in Figure 2. The out-of-pocket expenditure increased between 2008 (2.2%) and 2017 (2.6%). However, public expenditure decreased between 2009 (6.8%) and 2017 (6.3%).
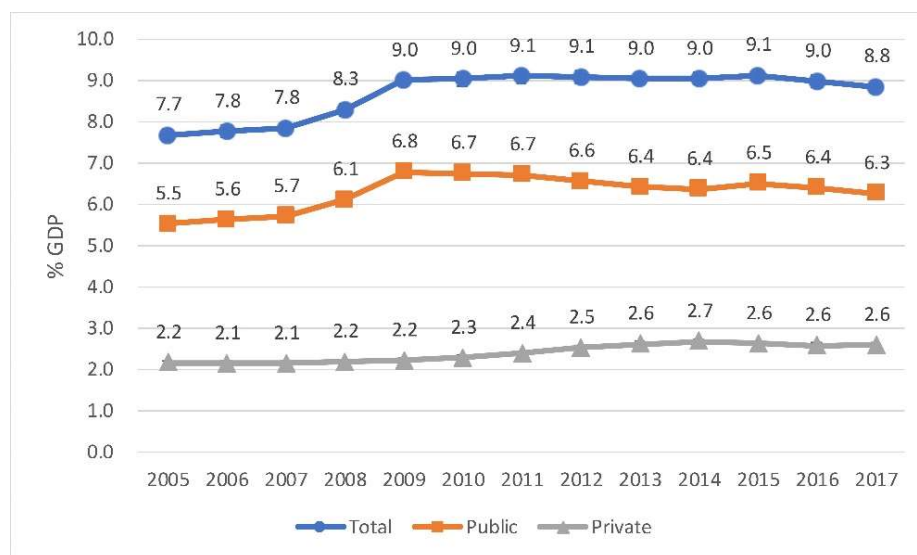

Figure 2. Spain. Health expenditure as a share of GDP. Source: Compiled by the authors based on Organization for Economic Co-operation and Development (OECD). Data extracted on 17 May 2019 10:07 UTC (GMT) from OECD. https://stats.oecd.org/.

In 2017, spending on healthcare in the European Union stood at 9.6% of gross domestic product, ranging from over 11% in France and Germany to less than 6% in Romania (Table 1). During the crisis, ten EU countries decreased health expenditure in real terms. Total health expenditure per capita in Spain during 2017 was 2446 euros, which is below the EU average of 2773 euros [9].

3.5. Out-of-Pocket Expenditure

Out-of-pocket payments are expenditures borne directly by a patient where neither public nor private insurance cover the full cost of the health good or service. They include cost-sharing and other expenditures paid directly by private households and should also in principle include estimations of informal payments to healthcare providers [9].

Across the EU, 18.2% of all health spending is borne directly by private households (Table 1). This figure ranges from around 10% in France, Luxembourg or the Netherlands to over 40% in Bulgaria, Latvia, and Cyprus. In Spain the share of total health spending financed by out-of-pocket payments is 23.8%.

3.6. Pharmaceutical Expenditure

After inpatient and outpatient care, pharmaceuticals represent the third largest component of health spending. In 2016, on average, medicines accounted for 16.8% of total health expenditure in EU countries (excluding medicines used in hospitals), but 19.1% in Spain (Table 1). The EU average of pharmaceuticals on a per capita basis was 416 euros, while in Spain it was over this amount (EUR 440) [9].

Table 2 shows the relevance expenditure in drugs has within the total public health expenditure in Spain. In 2017, 25% of the total public health spending was due to drugs.

Table 2. Public pharmaceutical expenditure as a share of National Health System 2017 [1].

Indicator	Spain
% Pharmaceutical expenditure (retail)	9.35
% Pharmaceutical expenditure (hospitals)	15.65
% Total pharmaceutical expenditure as a share of National Health System	25.00

[1] Out-of-pocket expenditure is excluded. Source: Compiled by the authors based on Public Health Expenditure and Pharmaceutical and Health Expenditure Statistics (Last updated: 22 April 2019). Ministry of Health and Ministry of Finances [12,13].

3.7. Avoidable Mortality (Preventable and Amenable)

Preventable mortality is defined as deaths that could be avoided through public health and prevention interventions, whereas amenable (or treatable) mortality is defined as deaths that could be avoided through effective and timely healthcare. The age-standardized rate of preventable mortality is lowest in Italy, Cyprus, and Spain, with rates at least 25% lower than the EU average. The age-standardized rate of amenable mortality is lowest in France, Spain, and the Netherlands, because these countries are among those with the lowest death rate from ischemic heart diseases and cerebrovascular diseases.

3.8. Share of Potentially Avoidable Hospital Admissions Due to Five Chronic Conditions

The share of potentially avoidable hospital admissions due to five chronic conditions (diabetes, hypertension, heart failure, chronic obstructive pulmonary disease (COPD), and bronchiectasis and asthma is a good indicator of the performance of the health system in managing chronic diseases. Spain's share (6.35%) is over the EU average (5.5%) (Table 1) [9].

3.9. Waiting Times for Elective Surgery

The data presented here focus on two high-volume procedures: cataract surgery (Figure 3) and hip replacement (Figure 4). In 2016, the average waiting time for people who underwent cataract surgery in Spain was around three months. The waiting time for cataract surgery increased from 83 days in 2009 and 2010 to 108.6 in 2016 [9].

In 2016, the average waiting time for people who underwent hip replacement surgery in Spain was around five months. The waiting time for hip replacement surgery increased from 117 days in 2005 to 158 in 2016. In 2011, the decreasing trend that begun in 2006 ceased and the waiting time started increasing [9].
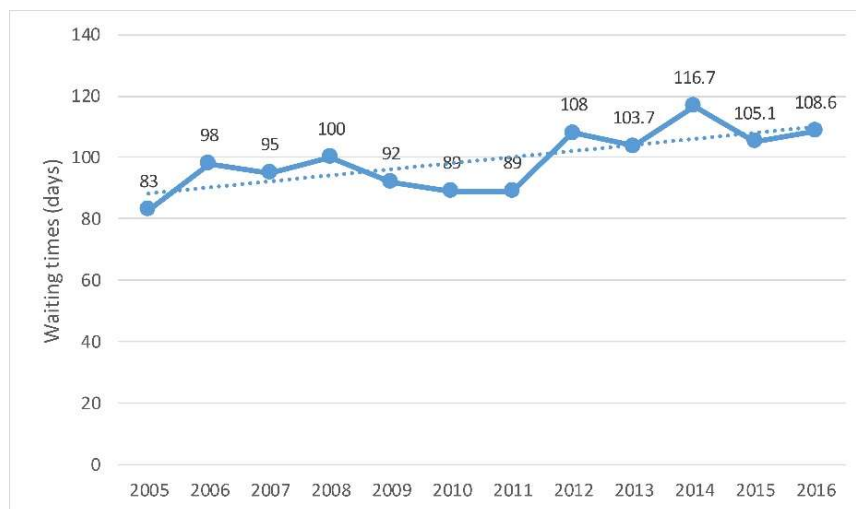
Figure 3. Spain. Waiting times of patients for cataract surgery, trends since 2005. Source: Compiled by the authors based on Organization for Economic Co-operation and Development (OECD) Health Statistics 2018. OECD/EU (2018), Health at a Glance: Europe 2018: State of Health in the EU Cycle [9].
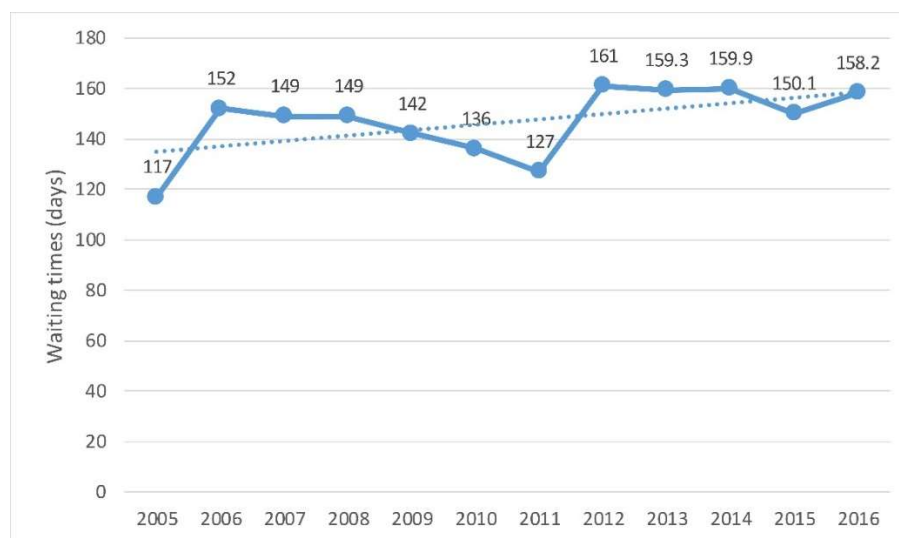

Figure 4. Waiting times of patients for hip replacement, trends since 2005. Source: Compiled by the authors based on OECD Health Statistics 2018. OECD/EU (2018), Health at a Glance: Europe 2018: State of Health in the EU Cycle [9].

3.10. Readiness of National Health Data Governance Frameworks to Support the Use of Electronic Health Record Data

Electronic Health Records (EHRs) are used by 100% of primary healthcare doctors in Spain, and according to the Ministry of Health data, in February 2019, 91.99% of prescriptions in Spain were e-prescriptions. E-prescribing allows doctors to write prescriptions that can be retrieved by a pharmacy electronically [14].

In 2016, the OECD Health Care Quality Indicators Expert Group conducted a study of 30 countries to explore the data governance and technical and operational factors that would support them in the development of national health information and research programs from data held within EHRs. The study explored several key indicators of the readiness of national health data governance frameworks to support the use of EHR data to fulfil national health information and research objectives. These indicators include: (1) having a legal framework that allows data within EHR systems to be extracted for statistical and research purposes, subject to suitable safeguards; (2) having a national

EHR plan or policy that includes the statistical and research uses of these data; (3) engaging in the extraction of data from EHR systems to create national datasets; and (4) analyzing data extracted from EHR systems for key national monitoring and research objectives. According to this study, Spain is far from the most advanced countries: Canada, Denmark, Finland, New Zealand, Singapore, Sweden, the United Kingdom (England and Scotland) and the United States; although it is ahead of countries such as Australia, France, and Japan [15].

4. Discussion

4.1. Main Challenges of the Spanish National Health System

The main challenges that the Spanish National Health System must face are the following:

1. The rising old age dependency ratio. This growing proportion of elderly people means an increasing demand of healthcare services due to a prevalence of chronic diseases that will cause an increase in health expenditure.
2. Health expenditure pressures and the likely increase of out-of-pocket expenditure.
3. Drug expenditures, both retail and consumed in hospitals.
4. Waiting lists for surgery.
5. The percentage of potentially preventable hospital admissions.
6. The use of EHR data to fulfil national health information and research objectives.

4.1.1. The Rising Old Age Dependency Ratio

The National Health System has a limited influence on the first of the challenges. The answer to these challenges lies in government policies in the areas of social services and labor conditions among others. Although the National Healthcare System cannot influence on this challenge, it can and must manage its consequences. The aging of the population produces higher morbidity and more chronic diseases, which cause a higher usage of health services. Therefore, the quality of attention that chronic patients receive must improve in order for them to improve their self- care and control, to avoid hospital stays, and to lower the usage of primary care centers.

4.1.2. Healthcare Expenditure Pressures and the Likely Increase of Out-of-Pocket Expenditure

Although the Independent Fiscal Authority expects an increased public health expenditure rate of 4% during the next few years, the uncertainty on the level of economic growth entails that public health expenditure is not likely to increase. However, the increase in population over 65 is very likely [16].

This challenge threatens the accessibility to healthcare services because the out-of-pocket expenditure increase coincides with the public health expenditure decrease. At an aggregate level, the share of out-of-pocket spending in total health spending reflects the degree of financial protection in a country along the three dimensions of coverage—the share of the population covered, the range of services included in a public benefit basket, and the proportion of costs covered by collective third-party payer schemes for each service [9].

4.1.3. Drug Expenditures, Both Retail and Consumed in Hospitals

Spending on medicines in Spain currently comprises 25% of the total health expenditure. The increase in drug expenditures is a complex phenomenon influenced by various related factors. The aging of the population together with the increased prevalence of chronic diseases and the increase in drug prices caused by the continuous introduction of new and more expensive drugs are factors that help explain it. However, others are just as important, such as clinicians' way of practicing medicine; the organization of the health system, which may have ways to control or influence drug prescription; the existence of guidelines; the degree of generic penetration in pharmaceutical consumption; and the work of pharmaceutical commissions, which analyze and decide on the incorporation of new drugs to

the drug formulary of hospitals. The drug industry has a notable effect on increased expenditures by regularly introducing new active principles to the market, some arguably without added value over existing ones and with a higher price [17].

More than 13,700 pharmaceutical products were in the pipeline in 2016, of which nearly 6900 were in a clinical phase of development. Oncology is by far the most targeted therapeutic area, representing almost one-third of the medicines in development in 2016, followed by prophylactic vaccines and anti-infectives, and antidiabetics [15].

4.1.4. Waiting Lists for Surgery and the Preventable Hospital Admissions Ratio Represent Two Quality Issues That Also Should Be Considered

Waiting times are the result of a complex interaction between the demand and supply of health services. The demand for elective surgery is determined by the health needs of the population, progress in medical and surgical technologies, and patient preferences. On the supply side, the availability of surgeons and other staff in surgical teams, as well as the supply of the required equipment and the occupation of beds in hospitals affect surgical activity rates [9]. Waiting lists are a long-standing problem that has proven to be very sensitive to fluctuations in the volume of public health expenditure. The increase in waiting lists in Spain since 2011 has coincided with the decrease in public spending (Figures 2–4). This situation reflects a weakness in the efficacy, efficiency, and quality of the health system.

Reducing the rate of potentially preventable hospital admissions, which in Spain is above the EU average, requires improving the care of people with chronic conditions. This is a very important task as the number of patients with chronic conditions is expected to rise. Better indicators about caring for chronic conditions are not available.

4.1.5. The Use of EHR Data to Fulfil National Health Information and Research Objectives

The foregoing leads to the need to improve the efficacy, efficiency, and quality of the health system. As noted above, IT brings improvements in efficiency and quality of operations management in health systems by reducing operating costs of clinical services, reducing administrative costs, and enabling entirely new modes of care. Health data constitutes a significant resource to improve population health and the effectiveness, safety, and patient-centeredness of healthcare systems. Spain does not have national legislation or regulation to allow extraction of data from the national EHR system for national research or statistical purposes. As a result, databases must be administered by each Health Authority (region). Spain also faces data governance barriers that limit researchers' access to electronic clinical data. These include a strong bureaucracy and a lack of written policies regarding how applicants may apply for access to data within EHR systems. Some IT providers are reluctant to provide access to data for research and/or apply financial charges for accessing data that may limit data accessibility [15,18].

All these circumstances make it necessary to develop a Big Data system for the National Health System of Spain.

4.2. A Big Data Solution for the National Health System of Spain

As listed along this article, the current challenges of the Spanish National Health System have been identified through a list of indicators obtained from the exhaustive analysis and monitoring of a heterogeneous set of data sources. Most of these sources are in fact provided by the National Health System, so it is in its hands to better understand and manage these data for the efficient use of the system's resources. This section applies the principles and criteria analyzed so far to draft a proposal on a Big Data System for the National Health System of Spain (BDNHS).

4.2.1. Big Data Principles

The idea of "data-driven innovation" (DDI) that refers to techniques and technologies for processing and analyzing large volumes of data, which are commonly known as "Big Data", is becoming an important resource that can lead to new knowledge, drive value creation, and foster new products,

processes, and markets. DDI is a source of economic growth and development through two distinctive "channels": [19]

1. The economic properties of data suggest that data is an infrastructure-like resource which in theory can be used by an unlimited number of users and for an unlimited number of purposes as an input to produce goods and services.
2. The value-creation mechanisms of data analytics, which include using data analytics to:

    *   gain insights (knowledge creation)
    *   automate decision-making.

The concept of Big Data, commonly misused, should only be applied to datasets which exceed the capacity of common software tools regarding volume, speed, and variety of data, among others. The relation of Big Data with data mining and knowledge management can enlarge the concept of Big Data to establish at least three fundamental characteristics of Big Data solutions:

*   The use of high-performance IT infrastructure regarding the capacity of information management in terms of volume, speed, and variety.
*   The extraction of information as a result of large-scale data processing.
*   The generation of knowledge from that information and its use for the strategic management of an organization.

The first goal when analyzing large datasets in health services is to provide the best possible information to those in charge of decision making: doctors, managers, and other health professionals. Health organizations must use data to reach their own goals as well as those of the people they are aiding. Therefore, data use must be oriented towards generating value for each patient and for the health system itself. This activity should also be directed towards anticipating the needs and preferences of patients and planning the services they could require, such as in the case of chronic patients who should receive personalized treatment.

The patient must always be the center of any health system and, as such, of the health information. When employing these tools, medical practitioners must be aware of the expected turnaround and of whether their patients are receiving the right care in the right moment. Managers of health organizations must also be empowered to make decisions by receiving the right information, so they can plan ahead, set targets, and make real time corrections [20]. Knowledge generated by health systems can be defined as an integrated entity in which science, IT, and health assistance services join forces to generate outputs derived from health services.

Data related to health systems may be classified in one of the following three categories:

*   Traditional: stored in clinical records and department files.
*   Data generated by patients and citizens: patient reported outcomes, social networks, devices, and individual sensors
*   Omics data: obtained from molecular biology.

Individual clinical information contained in EHRs has some limitations, as it does not include over the counter drugs; data on lifestyle habits, such as sleep, diet, or physical activity; environmental risk factors; and information from the perspective of the patient or his/her priorities. Health outcomes reported by patients could be used in clinical trials to monitor their efficacy and security. To make use of a resource that provides real world data it should incorporate resources generated by patients on social media, apps, sensors, wearables, direct services to consumers, polls, and questionnaires.

Human gene sequencing and the systematic analysis of other individual profiles are currently generating new data, known as omic data. Most EHRs can only include that type of data as a non-structured report, and it would therefore be hard to use in aggregated studies. Future compilation of multi-omic data will require standard representation formats, and the use of such data

in clinical practice will need support systems to clinical decisions and new ethical guides. Genetic and omic individual information plays a key role in the concept of precision medicine, as it is commonly known in the US, or personalized medicine, the term adopted in the European Union [21].

Precision medicine promises to provide the right treatment at the right time, every time, to the right person. It employs information on genes, proteins, and other pathologies of a patient, his/her lifestyle, prescribed drugs, health condition, and environmental factors in order to determine the diagnostic or treatment of the disease. The main field of application is oncology, because it can act on the changes that promote the growth, division, and dissemination of cancer cells. Precision Medicine has the potential to radically transform medicine and healthcare, because it consists on individual treatments for individual patients and promises to be more effective and safer than existing ones. Biomonitors to select patients are being analyzed for the use of new drugs, to adapt the administration of drugs, and to control adverse effects.

The health ecosystem is a concept even wider than the health system itself, since it consists of a set of entities that can become a patient-focused cluster or even a hub, if they take advantage of the synergies they produce. These entities are, among others, healthcare centers and services, universities, research centers, health industries, and the public administration. They generate and process health related information, and they must all be considered when it comes to the production of knowledge, which will improve the overall system.

The use of clinical data together with other large datasets provides a great opportunity for the improvement of the quality of health assistance, plus investigation, education, and epidemiological awareness. The aggregated data related to a disease or treatment enhances its effectiveness and quality and improves cost-control and efficiency. They can be especially useful to make decisions related to funding treatments and setting the cost of drugs. The link of biological data with clinical, environmental, and demographic data will improve epidemiological knowledge of chronic and infectious diseases.

An example of Big Data application in healthcare is population segmentation, by linking information from different sources, such as primary care, specialized care, and pharmacies, to develop projections of the likely evolution of a group of patients and plan their treatment in terms of demand, risk, resources, and costs. This process is especially interesting in the case of chronic patients, so they can get thorough tests related to certain diseases, which are a priority for healthcare organizations as they entail an important risk for these patients. Another example is the case of high frequency patients at primary care services: those that go 12 or more times per year. Big Data can help unwind the patterns of these visits and treat them in a more efficient way.

The incorporation of data from different sources will allow the advance towards personalized medicine by creating advanced systems for clinical decision making, thanks to the combination of clinical, genomics, and environmental information. It will also allow managers and public health systems to make more informed decisions. Lastly, it will introduce great changes in research, by studying data obtained from large pools instead of study groups. The end goal of these three processes is for clinical decision making to be predictive, personalized, preventive, and participative (Figure 5) [22].

Patient's phenotyping consists on identifying individuals which coincide with clinical investigative criteria. There are already methods in place for automated phenotype. When these systems are employed for selecting patients eligible for a clinical trial, the relevant characteristics are derived from the data of medical records which can be used for answering investigation lines. Since many countries have electronic records in their public services, new studies could be carried out by linking data of individuals from two or more sources. This must be done taking into consideration the privacy of patients [23].

For pharmaceuticals, there is an expectation that judicious use of "real-world data" (RWD) will make all steps in the development and use chain more effective and efficient, including research and development (R&D), regulatory decision making, health technology assessment (HTA), pricing and reimbursement decisions, and treatment. The increased opportunity to analyze RWD and the development in analytics methods will undoubtedly create many advantages for patients. Increased

focus on the importance of patient involvement and the widespread discussion on patient insight and engagement will complement the technical developments [18].
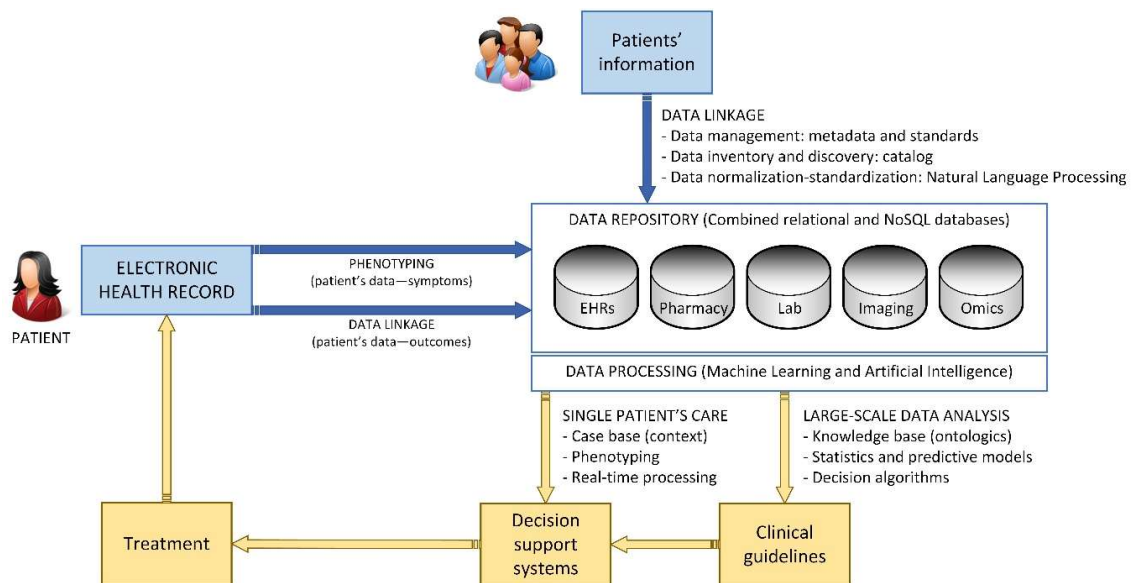

Figure 5. Healthcare knowledge management cycle related to single-patient care.

Big Data systems will allow real data evaluation and public health improvement, as well as considering social conditioners of health by gathering data from EHRs, which will progressively incorporate more information towards those goals [24].

In addition to the secondary use of patient data for clinical research and healthcare, there are other applications, most of which are halfway between EHRs and genetics. In this aspect, biorepositories and biobanks play a key role in generating data for a potential secondary use. Data repositories originating from clinical attention are accepted as a complimentary source of randomized clinical trials. This way, clinical research extends to real world evidence, which evaluates effectiveness and security during common clinical practice.

The future aggregation of omics and environmental data will require standard formats, support systems for clinical decisions, and new ethical and legal guidelines.

4.2.2. Data Acquisition

The main challenge in a project which employs large health related datasets consists of determining which are that data, how they can be intertwined, and what is their value. This decision must be made in a structured and unanimous way, and with explicit and objective criteria on quality, accessibility, and structure. Organizations must deal with issues such as integrating structured and quantitative data with non-structured qualitative sources, determining the data provenance, and the fragmentation of digital health records and proprietary systems, which makes interoperability more difficult.

The process of analyzing information begins with the selection of necessary data. This data corresponds to multiple areas, so they will be stored in different systems and databases, each of which will have its own structure. Before entering the system, this data must undergo a quality control that determines:

- If the data is compatible with the system (format and compatibility).
- If the data responds to existing needs (suitability and appropriateness).
- If the data is complete (completeness).
- If the data is valid (integrity and consistency).
- If the data is detailed enough (precision and accuracy).

- If the data is structured in a way that the system can process (standardization and normalization).

Data meeting the requirements of quality control can then undergo a pre-processing phase that may include modification of its original structure, indexing to stablish existing relations between them, and final storage in the repository of the system (Figure 6).
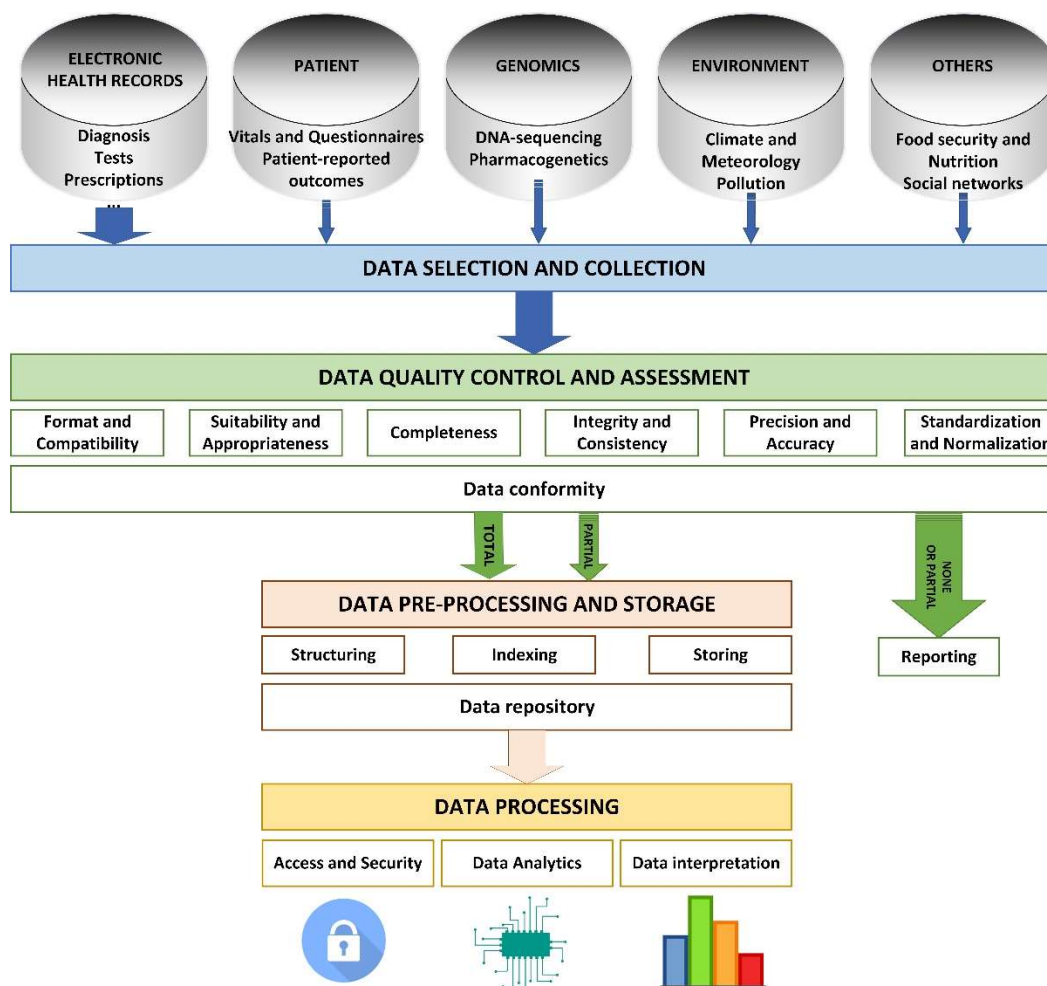

Figure 6. Phases for managing information processes in a system based on Big Data.

Lastly, users performing data analysis can access the data safely, undertake required operations, and interpret the results of the process, which generates new knowledge. The access to health records must be conceded based on the purpose of using the data rather than on the type of user. The model proposed by health systems entails that the data registered during assistance helps improve the process and obtain greater information on diagnosis and therapeutic options. Investigators will obtain real world data, which will improve the execution of clinical trials and make them more pragmatic and adaptative. Therefore, these trials will study, not just the efficacy, but also the effectiveness of new therapies (Figure 6).

One of the main aspects of information systems is the design of datasets and metadata, because it has a large influence on the entire storage, organization, and analysis of information. The structure of a database must bring the right level of detail for every data. The model of data in a project such as EHRs must consider the combinations of structured and non-structured data and find a better way to incorporate these in an efficient and effective way. Therefore, this aspect leads to the combined use of NoSQL databases, and of tools such as natural language processing [25]. These are especially relevant for structuring data originating from EHRs. Another relevant aspect is the fact that tools for Big Data

analysis are so specific that each project requires a combination of specific technological tools that once integrated give a global response to all needs of the system. For instance, a critical or time-sensitive system must resort to real time technologies.

One of the first steps in these projects is to make an inventory of the different datasets available with a detailed description by using metadata, which includes codes or standards which must be employed to guarantee semantic interoperability. Widely used standards are especially relevant in the case of federated systems because their managers must decide if the data is combined or stays separate.

Data safe havens are focused on trust, reliability, fulfilling legal and ethical criteria, and the management and publishing of data within socially acceptable limits. The criteria can also be related to trust, integrity, and availability, as well as an adequate and safe access point. Data safe havens have well defined governance systems, their safety is trusted, and the result of its investigations is transparent.

During the initial phases of the project the independent character of the regulatory body of the project must be established, which extends its control over the usage of the data by the public health system. This body must arrange the requirements for joining and signing the agreements with the entities that agree to share the data once it is decided that they fulfil the requirements.

4.2.3. Treatment of Data

Data governance can be defined as the management of data in terms of accessibility, availability, usage, integrity, and quality. The key is to be able to easily store data in available repositories, but also with the appropriate technical and government controls, auditable and audited in an effective way, and deemed trustworthy to all interested parties. The approval and kick-off of Data Health Systems requires an explicit compromise by the stakeholders of the national health system and must be controlled by a committee or specific working group.

The application of Big Data to the analysis of large datasets is based on two principles: first, health centers must have interoperable EHR, and second, the metadata must be made available by health systems. They must have a series of technological, semantic, and organizational standards that allow the exchange of health information keeping the coherence of meaning of the data.

The logic architecture of the system, that is, the conceptual scheme that describes its components and the relations between them must be coherent with the established governance of the data.

Depending on the final structure of the system, there are two main types of architecture:

- Centralized: it comprises a central node that works as the sole repository of the system data. This requires implementing a process for storing new data as they are generated, so that they are available for consultation and analysis by users of the system.
- Federated: each provider data center is at the same time the repository of its own dataset, which it only shares with the rest of the nodes of the system on demand, that is, when that data is required for analysis.

No model is better than the other. Instead, their advantages and disadvantages must be examined on a case by case basis. The main requirements regarding the adoption of standards and the capacity of growth and evolution must be applied on equal terms to centralized and federated architectures so that timely access to data with the required format is guaranteed. For instance, real time applications will benefit from centralized architecture, because it lowers the response time since all data compilation is done in advance. However, federated architecture can be more interesting for projects involving more than one organization, so that they can protect their own information and have absolute certainty that only what it strictly necessary will be shared.

A mixed model can also be adopted, which uses one method or another depending on the nature of each case. This possibility combines using a centralized repository for data whose analysis might be time sensitive, such as consulting a support system for making clinical decisions in the midst of a medical consultation, with using federated databases for information whose analysis does not require a quick execution, such as massive analysis of data for a clinical investigation project.

A data health dictionary containing a description of the archives available on the database must also be created. This will become a fundamental tool for anyone interested in accessing the content of the database. With this instrument, users can search the databases, filtering categories or record types.

The expected workflow of Data Health Systems anticipates that users with a research question will search on the data dictionary the availability of the data they require. After that, they must complete the research protocol on one of the available forms of the system. The next step is to obtain approval by the governance units to ensure the study is adequate and compliant with the existing regulation. Afterwards, the technical teams must prepare the data processing and its mapping with standard codes. Once the data is ready and its quality has been checked, it must be analyzed, and the results made available. If anonymized data contains any form of tracking that allows for its identification, there must be a frame for minimizing that risk.

All protocols for accessing the system must be easy to understand, available online, and transparent. Organizations as well as individuals must be certified and prove that they have the technical means to guarantee the safety of data. Any strategy to reduce risk according to the secondary use of data must be based on the improvement of the capacities of the users of the systems, and not on making the access more difficult with more complex and restrictive protocols.

There must be a balance between the best possible design of the studies and the practical aspects which make it possible, such as the data load and the costs associated to the participants.

Lastly, it is fundamental that each actor, including both institutions and individuals, may access and analyze data in a safe, effective, and efficient manner.

4.2.4. Data Interoperability

As mentioned, data homogenization is a key process to achieve the integration of the different sources and the semantic interoperability of the involved information systems [26]. The use of extended standards to represent and manage both captured and produced datasets will allow proper data analysis, provide a wider set of tools to process information and support further reuse of the outcomes. This way, the standardization of the data flow, at all levels of the health system, is of utmost importance to produce valuable results.

Standards in data management related to health systems apply to three different levels:

- Messaging standards (for data communication). These standards define the way the different IT systems communicate with each other, providing a set of rules that guarantees that the exchanged information is solid. These rules set the format, structures, and data types that build each of the allowed messages to be shared among the tools within the system and to communicate with external ones. Digital Imaging and Communications in Medicine (DICOM) [27] for digital imaging or Health Level Seven (HL7) [28] are two of the most used ones in worldwide health systems interoperability.

- Documentation standards (for human communication). If messages are intended for digital systems, documents structure information to be managed by humans. In a health system, everyday hundreds of documents are filled and read by professionals. These documents reflect, among others, valuable information about the patients and the processes they are subjected to. Standards here define common structures to capture and present information, from and to the system users. In turn, these structures enable software systems to analyze and process the information, build useful messages, and improve interoperability. Clinical Document Architecture (CDA®) [29], Continuity of Care Record (CCR) [30], and Continuity of Care Document (CCD) [31] are extended clinical documentation standards.

- Terminology standards (to represent information). This set of standards provides a common lexis to define, represent, and code medical concepts, such diagnosis, disease, medication, etc. This health language will be later used to fill the clinical documents shared by the whole health system. Systematized Nomenclature of Medicine – Clinical Terms (SNOMED CT) [32] or 10th

revision of the International Statistical Classification of Diseases and Related Health Problems, (ICD-10) [33] (for disease statistics) are two of the most relevant in this area.

The proper selection of the set of standards implemented to manage the information among all the different levels of the health system will directly impact its interoperability, as well as the functionalities offered to its users.

4.2.5. Bioethical and Legal Aspects

From an ethical point of view, quite a few similarities to the legal field can be observed. The fact that IT is going to play an increasingly important role in health systems seems to be widely accepted, since its potential as a key instrument for the transformation of the current model is appreciated. Table 3 summarizes the key aspects that must be considered for a bioethical discussion [34].

Table 3. Key points to be considered in a bioethics discussion.

Precision and Validity of Data and Algorithms
Consent to data sharing and new ways to obtain them, such as automated sharing, voluntary communication, amplified consent, or dynamic consent.
Confidentiality, which deals with the repercussion of coding or pseudo anonymization, and the anonymization of data in the analysis of information, as well as the dilemma between right to privacy and public interest.

Source: León P. Key Points for an Ethical Evaluation of Healthcare Big Data [30].

Some ethical challenges are straightforward and need to be guarded against, such as concerns that algorithms may mirror human biases in decision making. Others, such as the possibility for algorithms to become the repository of the collective medical mind, have less obvious risks but raise broader ethical concerns [34,35].

Patient's rights must include the principles of respect to people, human rights, participation of those with morally relevant interests, and the responsibility of the decisions taken. Precision and validity of data and algorithms are fundamental to making the right choice.

Informed consent must be explicit and refer to the decision made on the information flow. Specific consent must be granted for data processing, although secondary analysis can be performed without an extra consent if the United Nations Educational, Scientific and Cultural Organization (UNESCO) principles are respected: appropriate legal background, evaluation by the ethical and research committee, public interest of the investigation, and impossibility of a new consent [36].

The right to privacy and public interest are intertwined. There are private interests in obtaining public interest and public interest in privacy protection, which encourages cooperation. They must both come to an agreement. In fact, the new European General Data Protection Regulation states that it might be necessary to obtain data without consent for public interest reasons. The law allows for pseudonymization, but it is hard to assure that anonymity can be preserved if different databases of different provenance are combined.

The responsibility of health professionals and managers of the health system, which extends to institutions, covers appropriate data analysis including its goals, quality, precision, and statistical procedure.

The Advisory Panel of Experts on Health Information Infrastructure (OECD Health Committee) identified key data governance mechanisms supporting privacy-protective data use [6]:

1. The health information system supports the monitoring and improvement of healthcare quality and system performance, as well as research innovations for better healthcare and outcomes.
2. The processing and the secondary use of data for public health, research, and statistical purposes are permitted, subject to safeguards specified in the legislative framework for data protection.
3. The public are consulted upon and informed about the collection and processing of personal health data.

4. A certification/accreditation process for the processing of health data for research and statistics is implemented.
5. The project approval process is fair and transparent and decision making is supported by an independent, multidisciplinary project review body.
6. Best practices in data de-identification are applied to protect patient data privacy.
7. Best practices in data security and management are applied to reduce reidentification and breach risks.
8. Governance mechanisms are periodically reviewed at an international level to maximize societal benefits and minimize societal risks as new data sources and new technologies are introduced.

4.3. Proposal for the Learning Health System of Spain

Although the essential principles of the National Health System of Spain are universal coverage and public funding of the system via taxes, it is also defined as an integrated health system which provides both primary as well as specialized care. This task is undertaken by public employees in public institutions, and under the management of regional health systems.

The information services of the health system reflect this decentralized nature, although the Ministry of Health is responsible for the systems and information technologies shared by all health services. An Inter-territorial Board of the National Health System in which all the regions that take part oversee the coordination between the healthcare providers. It is tasked with acquiring knowledge, debating, and advising on the essential functions of the National Health System, and planning, coordinating, and evaluating its performance.

All healthcare services in Spain have EHRs. They add efficacy, efficiency, and safety to the health attention process. They register essential data such as age, sex, diagnoses and procedures, prescribed drugs, and the results of explorations, among others. All regions register public health records, prescription drug usage, and labor and environmental health. The EHR also provides information on adverse medication effects among other relevant registries for the Health System. All this information helps build a system which allows the analysis of all these sources to benefit patients, the health system, researchers, and society.

4.3.1. Data Acquisition and Traceability

According to the report on the implementation of the interoperable EHR of the National Health System, in 2017 36,193,454 citizens had references with available contents, which accounts for 77.88% of the population [37].

Despite that, the system aims to make medical records accessible to patients and health professionals in the assistance process. The system is not conceived for investigation purposes, but for generating assistance documents on demand and in relation to specific patients.

In order to use information for research purposes it would be necessary to implement a more detailed data structure, allowing for more precise information management and selection, such as the selection of specific groups of patients; and apply a number of measures to guarantee safe access to that data, such as anonymization in accordance with existing laws covering clinical investigation. It would also be convenient to complete the information with other types of non-clinical data, such as environmental, which could be potentially relevant in a research project.

The information contained in medical records that can be found in EHRs of healthcare systems should be updated, and it should register the information generated by patients. The use of that data requires an architecture backed by structured and coded data. These applications require merging data from different sources. The result is a specific database which can be accessed via safe access when necessary.

4.3.2. Architecture Overview

Most of the main functionalities and building blocks of the presented Spanish Learning Health System have already been described in previous sections. This part summarizes the authors' approach into a holistic technical architecture that would guide a preliminary implementation for the introduced solution. This architecture is composed of three main layers, each of them capable of being implemented in a centralized, distributed or combined manner, addressing three different stages on data gathering, data processing, and data exploitation.

- Layer 1: Data Gathering and Data Homogenization

Big Data principles show the relevance of standardized data capturing and the different processes needed to guarantee quality, consistency, and compatibility. This layer implements all required connectors to the different heterogeneous data sources that filter and convert information, providing layer two with homogeneous and standard datasets. Most of the integration capabilities of the final solution will rely on the connectors set here.

- Layer 2: Big Data and Machine Learning

The core of the proposed system will be composed by two main blocks: the Big Data tools, that classify, organize, and store all captured data; and the machine learning environment implemented, which includes the algorithms and neural networks, fed by the Big Data tools, that train the whole system and make it "learn". This layer will produce most of the relevant outcomes of the Learning System.

- Layer 3: Presentation and Data Analysis

Finally, the presentation layer will expose the processed results, as well as all captured and classified health information. Through different data warehouse solutions, clinic workstations, and further connections to specialized data analysis tools, these results will generate and present new valuable medical knowledge to improve the efficiency of the whole Spanish National Health System.

4.3.3. Road Map

The main stages for implementing the BDNHS are summarized in Table 4.

The project must have full institutional backing to guarantee its acceptance and the participation of all health services, universities, research centers, and scientific societies.

Other than being explicit, the backing must also translate into awarding the necessary budget at least for the first two years of the project. Project managers must also look into funding sources, which should add to institutional funding additional sources such as European Union grants.

A project of this nature must count on a team of professionals coming from different disciplines such as clinicians, researchers, managers, health system experts, health IT experts, lawyers, and bioethics experts.

The project manager must respond directly on the development of the project to the Ministry and the Interregional Council. The project team must have Ministry and regional representatives, and it would be recommended to have other entities of the health system take part, such as professional associations, scientific societies, IT company clusters, research and teaching centers, and patient associations.

The operational targets require defining the different applications that the BDNHS intends to cover, because its deployment must be compatible with clinical assistance, and health services' research operations and day to day management. These environments have different needs and procedures, so they must compromise on solutions that allow the harmonization and tend to the requirements of them all.

Table 4. Road map for a Big Data System of the Health System of Spain.

Step	Phase	Description
1	Approval by institutions	The project must be approved by the Health Ministry and the Interterritorial Board, who must explicitly back it.
2	Budget funding	The project must be awarded a budget. It will be covered with Ministry and regional funds and other sources, such as EU funds.
3	Project team selection	The team will be comprised of scientific groups, patients, and research centers, among others.
4	Operational goals definition	The following goals must be considered: clinical assistance, investigation, and management of health services and systems.
5	Digital health platform creation	An online bio health center dedicated to digital health must be created
6	Legal and institutional frame	Government units, task units, procurement, agreements, and partnerships. Ministry and Interterritorial Board accountability. Data protection rules, biomedical research, intellectual property.
7	Technical requirements definition	Definition of data modeling. Data protection rules. System architecture model
8	Information security plan	System Access control. System Access registry. Information use control. Registry of information. Access, auditing of system access and data use. Information integrity. Information availability
9	Technical proposal	Technical proposal that details the necessary steps for system implementation
10	Management	Management groups, with Ministry of Health and regional representatives. Usage rules. Protocols to access and use data
11	Pilot execution	Required to optimize developed tools
12	Global evaluation	Use of the system and impact on efficacy, effectiveness, and efficiency of the National Health System, which should impact the health of the population

In order to obtain results in the short-term realistic milestones must be set. This would allow all parties to quickly profit from the effort and investment undertaken, evaluate the impact of the measures under implementation, reinforce trust and compromise among clinicians, bring benefits to patients, and overall, add value to the project.

There is currently no net devoted to digital health, nor a reference center in terms of health IT research in Spain. One of the critical factors for the success of the BDNHS is to count on a critical mass of experts in the integrated management of, among others, clinical, socioeconomic, and environmental data. Therefore, it is critical to create a net dedicated to IT health, as well as educating specialists in the field.

There are two areas that fall under legal analysis in this project. The first one refers to the legal nature of the BDNHS management unit, because it should not be part of the basic public administration structure. Instead, it should be an entity with autonomous management capacity, although closely related and controlled by the Health Ministry and Health Services. On the other hand, a thorough analysis of existing legislation covering data protection must be undertaken to consider the requirements and restrictions on information usage. Another relevant aspect is the legal frame that governs clinical investigations, which could require changes depending on the applications and mechanisms defined on the BDNHS.

Intellectual property must also be examined regarding the results of the BDNHS. Participating institutions must undertake a large investment. Their goal is that the resulting system works as fundamental support to the transformation, improvement, and sustainability of the current health system. However, an additional result could be the offering of services to third parties, such as private healthcare providers, research centers or health-tech companies, among others. All proposals to modify existing patient rights and data protection must consider the bioethical aspects of data treatment.

Among the requirements for the system, the following three aspects stand out:

1. The definition of the BDNHS data models, considering the diversity of data that will be used; the information sources that will provide that data; and the use of that data inside the BDNHS.
2. The standards policy on data management: acquisition, standardization, storage, exchange, and processing, among others.
3. The election of a model of architecture for the system, which could be federated, centralized or hybrid. However, the most appropriate final architecture could end up being a hybrid model, with a centralized repository for clinical information, whose data structure would have a high degree of similarity, and federated repository for nonclinical information, which would cover diverse data originating in different sources.

Health data has the largest available protection from a legal point of view in the EU. Therefore, an essential aspect of the BDNHS project must be the definition and application of a security plan for the information that guarantees upholding existing rules. This plan must define measures related to the control and registry of system access and information use, auditing of system access, and use of data, and integrity of information through data quality control and security copy mechanisms. It must also assure the availability of information through mechanisms of redundancy of backing measures that can be quickly activated in case of partial or complete failure of the main system.

All previously mentioned technical aspects must be integrated in a technical proposal that details the existing steps necessary to implement the BDNHS.

From an organizational and institutional point of view, the procedures that are needed to govern and manage the system resulting from the project must be defined once it is implemented and put to work. The governing bodies must be defined, as well as its internal rules and the protocols that will govern the access and use of existing data.

It is also necessary to run a pilot in a defined and controlled environment with an adequate size that allows obtaining conclusions on the functioning of the BDNHS. Every time an element or component is introduced managers must take into consideration the need to run a pilot, and, if required, its reach. A new element or component must be defined as a type of data, an information source, a use of that data, or a region, among others.

One of the key aspects for the BDNHS to add value is the evaluation of the results obtained from the project. The main elements to take into account are the use of the system and its impact on the efficacy, quality, efficiency, and effectivity of the NHS, which must translate into an impact on the health of the population. The evaluation must take place periodically, with the goal of checking the evolution of these parameters.

4.3.4. Critical Success Elements

The following elements are considered critical for the success of the project:

1. The leadership of the Ministry of Health, which will allow the resistance that the regional health service could present to be overcome, by assigning a budget and handling the legal requirements for the creation of the bodies of the project, as well as modifying data protection rules and health research if necessary, and taking care of intellectual property matters.
2. The agreement of the Interregional Healthcare Council creating the BDNHS, explicitly backing it and taking over the compromises required: data provision, and resource and budget allocation.
3. The granting of a budget allowance for at least the first two years of execution of the project.
4. The creation of a management body for the project, whose team must be comprised of renowned professionals with leadership skills. The director must have an executive profile but with flexibility and negotiating skills to reach agreements between the stakeholders of the project.
5. The implication of all the Regional Health Services that comprise the National Health System of Spain, who must obtain value in the short term in order to promote and reward their participation.
6. The training of professionals of the participating entities in IT health in order to have the best available human resources to execute the project.

7. The creation of a management unit that will extend its control over data usage to improve public health as well as the health system. This management unit must determine the standards for joining the program. It must also sign the required agreements with the entities that decide to share their data once it is accredited that they comply with the requirements. This unit must be of permanent nature.

8. The choice of an architecture model that considers the organization, technical, and budget conditions of the participants, especially those of the 18 regional health services.

9. The consideration and thorough application of the recommendations provided by bioethics experts to provide trust to citizens and professional entities as well as patient associations.

10. The compliance with data protection rules and the review of existing Spanish legislation in order to take advantage of the opportunities that it brings for projects such as the BDNHS.

11. The participation of patients in the project. This requires a communication plan as well as making the project known to the broader society. The plan must convey trust regarding citizen's privacy in line with the previous two recommendations. Trust requires transparency regarding data sources, their processing, usage, and results.

12. The achievement of short-term milestones, bringing early benefits to users, and providing returns and visibility to the project.

5. Conclusions

The Spanish National Health System is facing various demographic, economic, and management challenges that will force it to undergo a deep transformation of the existing model. These challenges directly impact the healthcare quality and expenditure, causing an inefficient use of the overall system resources. In turn, each of these challenges are linked to several well-identified indicators obtained from the analysis of different heterogeneous datasets. In this sense, the awareness and traceability of these indicators, as well as their control (when possible), will improve the efficiency and efficacy of the whole health system, while also enhancing the quality of the services offered. One of the key aspects for the success of this process is the creation of a Learning Health System which should optimize knowledge generation and its capitalization by applying it to healthcare and health system management.

IT enablers are a fundamental tool to address the complexity and vastness of health data as well as the urgency that clinical and management decisions require. Big Data solutions are a perfect match for that problem in health systems.

The creation of the BDNHS entails following these requirements:

- Identifying the required data and their sources.
- Implementing a repository of data to compile it before its analysis.
- Defining an architecture model according to the needs of the health system.
- Incorporating data protection measures in accordance with existing rules and applicable bioethics guidelines.
- Using the most adequate data tools for the defined system model.
- Creating the management unit required for putting the model to work and managing the system, with participation from regional health services, strong institutional backing, and the required budget allocation.
- Complying with bioethical legal and data transparency guidelines, and data processing and use.

In summary, the Spanish National Health System is under pressure to undertake major reforms. For such purpose it can, and should, make use of existing technological tools which can serve as a starting point to face these challenges, even though they are not enough to solve them all, as a means to start the project that will result in a Learning Health System.

References

1. Muir Gray, J.A. *How to Get Better Value Healthcare*, 2nd ed.; Oxford Press Ltd.: Oxford, UK, 2011.
2. Porter, M.E.; Lee, T.H. The Strategy That Will Fix Health Care. *Harvard Business Review*. October 2013, pp. 50–67. Available online: https://hbr.org/2013/10/the-strategy-that-will-fix-health-care (accessed on 6 September 2019).
3. Tremolada, M.; Schiavo, S.; Varotto, S.; Basso, G.; Pillon, M. Patient Satisfaction in Italian Childhood Cancer Survivors: Human Aspects of Treatment as a Key Factor in Patients' Quality of Life. *Health Soc. Work* **2015**, *40*, e148–e155. [CrossRef]
4. OECD Improving Health Sector Efficiency. *The Role of Information and Communication Technologies*; OECD Publishing: Paris, France, 2010.
5. OECD. *Strengthening Health Information Infrastructure for Health Care Quality Governance: Good Practices, New Opportunities and Data Privacy Protection Challenges*; OECD Health Policy Studies; OECD Publishing: Paris, France, 2013. [CrossRef]
6. OECD. *Health Data Governance: Privacy, Monitoring and Research*; OECD Health Policy Studies; OECD Publishing: Paris, France, 2015. [CrossRef]
7. Bernstein, J.; Friedman, C.; Jacobson, P.; Rubin, J. Ensuring Public Health's Future in a National-Scale Learning Health System. *Am. J. Prev. Med.* **2015**, *48*, 480–487. [CrossRef] [PubMed]
8. WHO. *Health Systems: Improving Performance*; The World Health Report; World Health Organization: Geneva, Switzerland, 2000; Available online: https://www.who.int/healthinfo/paper30.pdf (accessed on 3 June 2019).
9. OECD/EU. *Health at a Glance: Europe 2018: State of Health in the EU Cycle*; OECD Publishing: Paris, France; EU: Brussels, Belgium, 2018. [CrossRef]
10. Ministerio de Sanidad, Consumo y Bienestar Social. Informe anual del Sistema Nacional de Salud. (Ministry of Health. Annual Report of National Health System). 2017. Available online: https://www.mscbs.gob.es/estadEstudios/estadisticas/sisInfSanSNS/tablasEstadisticas/InfAnSNS.htm (accessed on 3 June 2019).
11. Instituto Nacional de Estadística. Indicadores Demográficos Básicos. (Spanish National Statistics Institute. Basic Demographic Indicators). Available online: http://www.ine.es/dynt3/inebase/es/index.htm?padre=2077&capsel=2081 (accessed on 6 September 2019).
12. Ministerio de Sanidad. Estadística de Gasto Sanitario Público. (Ministry of Health. Public Health Expenditure Statistic). Available online: https://www.mscbs.gob.es/estadEstudios/estadisticas/inforRecopilaciones/gastoSanitario2005/home.htm (accessed on 6 September 2019).
13. Ministerio de Hacienda. Estadística de Gasto Farmacéutico Y Sanitario (Ministry of Finance Pharmaceutical and Health Expenditure Statistics). 2019. Available online: http://www.hacienda.gob.es/es-ES/CDI/Paginas/EstabilidadPresupuestaria/InformacionAAPPs/Indicadores-sobre-Gasto-Farmacéutico-y-Sanitario.aspx (accessed on 6 September 2019).
14. Ministerio de Sanidad. Implantación de la Receta Electrónica en el Sistema Nacional de Salud (Ministry of Health. Implementation of Electronic Prescription in National Health System). Available online: http://www.mscbs.gob.es/profesionales/recetaElectronicaSNS/NIVEL_DE_IMPLANTACION_febrero_2019.pdf (accessed on 6 September 2019).
15. OECD. *New Health Technologies: Managing Access, Value and Sustainability*; OECD Publishing: Paris, France, 2017. [CrossRef]
16. Informe Sobre la Actualización del Programa de Estabilidad 2019–2022. Informe 32/19. AIREF. Autoridad Independiente de Responsabilidad Fiscal 2019 (Report about the Update of the Stability Programme. Report 32/19. AIREF. Independent Authority for Fiscal Responsibility). Available online: https://www.airef.es/wp-content/uploads/2019/05/informe-ape/Informe_AIReF_APE-2019-2022.pdf (accessed on 6 September 2019).
17. Carnicero, J. La Prestación Farmacéutica en el Sistema Nacional de Salud y en la Comunidad Foral de Navarra. In *Pharmaceutical provision in National Health System and in the Autonomous Community of Navarra*; Gobierno de Navarra, Departamento de Salud: Pamplona, Spain, 1996.

18. Eichler, H.G.; Bloechl-Daum, B.; Broich, K.; Kyrle, P.A.; Oderkirk, J.; Rasi, G.; Wenzl, M. Data rich, information poor: Can we use electronic health records to create a learning healthcare system for pharmaceuticals? *Clin. Pharmacol. Ther.* **2019**, *105*, 912–922. [CrossRef] [PubMed]

19. OECD. *Data-Driven Innovation: Big Data for Growth and Well-Being*; OECD Publishing: Paris, France, 2015. [CrossRef]

20. Rojas, D.; Carnicero, J. Big Data and Public Health Systems: Issues and Opportunities. *Int. J. Interact. Multimed. Artif. Intell.* **2018**, *4*, 53–59. [CrossRef]

21. Weng, C.; Kahn, M.G. Clinical Research Informatics for Big Data and Precision Medicine. *Yearb. Med. Inform.* **2016**, *10*, 211–218.

22. Martin-Sanchez, F.; Verspoor, K. Big data in medicine is driving big changes. *Yearb. Med. Inform.* **2014**, *9*, 14–20. [PubMed]

23. Carnicero, J.; Rojas, D. Healthcare Decision-Making Support Based on the Application of Big Data to Electronic Medical Records: A Knowledge Management Cycle. In *Leveraging Biomedical and Healthcare—Semantics, Analytics and Knowledge*; Kobeissy, F., Wang, K., Zaraket, F., Alawieh, A., Eds.; Elsevier: San Diego, CA, USA, 2018.

24. Ahmed, M.; Ullah, A.S.B. Infrequent pattern mining in smart healthcare environment using data summarization. *J. Supercomput.* **2018**, *74*, 5041–5059. [CrossRef]

25. Modoni, G.E.; Sacco, M.; Terkaj, W. A survey of RDF store solutions. In Proceedings of the 2014 International Conference on Engineering, Technology and Innovation (ICE), Bergamo, Italy, 23–25 June 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 1–7.

26. Modoni, G.E.; Veniero, M.; Sacco, M. Semantic knowledge management and integration services for AAL. In *Italian Forum of Ambient Assisted Living*; Springer: Cham, Switzerland, 2016; pp. 287–299.

27. DICOM Standard. DICOMWeb™ Publications. Available online: https://www.dicomstandard.org/current/ (accessed on 22 June 2019).

28. Introduction to HL7 Standards. Health Level Seven (HL7) International. Available online: http://www.hl7.org/implement/standards/index.cfm?ref=nav (accessed on 22 June 2019).

29. Clinical Document Architecture, Release 2 Health Level Seven (HL7) International. Available online: https://www.hl7.org/implement/standards/product_brief.cfm?product_id=7 (accessed on 22 June 2019).

30. Standard Specification for Continuity of Care Record (CCR). Active Standard ASTM E2369. Available online: https://www.astm.org/Standards/E2369.htm (accessed on 22 June 2019).

31. Continuity of Care Document (CCD): Changing the Landscape of Healthcare Information Exchange. Corepoint Health. Available online: https://corepointhealth.com/resource-center/white-papers/continuity-care-document-ccd-changing-landscape-healthcare-information/continuity-care-document-ccd-changing-landscape-healthcare-information-exchange/ (accessed on 22 June 2019).

32. SNOMED CT: 5-Step Briefing. © SNOMED International 2019. Available online: http://www.snomed.org/snomed-ct/five-step-briefing (accessed on 22 June 2019).

33. International Classification of Diseases (ICD). © World Health Organization. Available online: https://www.who.int/classifications/icd/factsheet/en/ (accessed on 22 June 2019).

34. León-Sanz, P. Key Points for an Ethical Evaluation of Healthcare Big Data. *Processes* **2019**, *7*, 493. [CrossRef]

35. Char, D.S.; Shah, N.H.; Magnus, D. Implementing machine learning in health care—Addressing ethical challenges. *N. Engl. J. Med.* **2018**, *378*, 981–983. [CrossRef] [PubMed]

36. UNESCO. *Report of the International Bioethics Committee of UNESCO on Big Data and Health*; UNESCO: Paris, France, 2017; Available online: https://unesdoc.unesco.org/ark:/48223/pf0000248724 (accessed on 3 June 2019).

37. Ministerio de Sanidad, Consumo y Bienestar Social. Proyecto de Historia Clínica Digital del Sistema Nacional de Salud. Informe de situación. Available online: http://www.mscbs.gob.es/profesionales/hcdsns/contenidoDoc/HCDSNS_Informe_de_situacion_20190401.pdf (accessed on 3 June 2019).