Deliverable D2.2

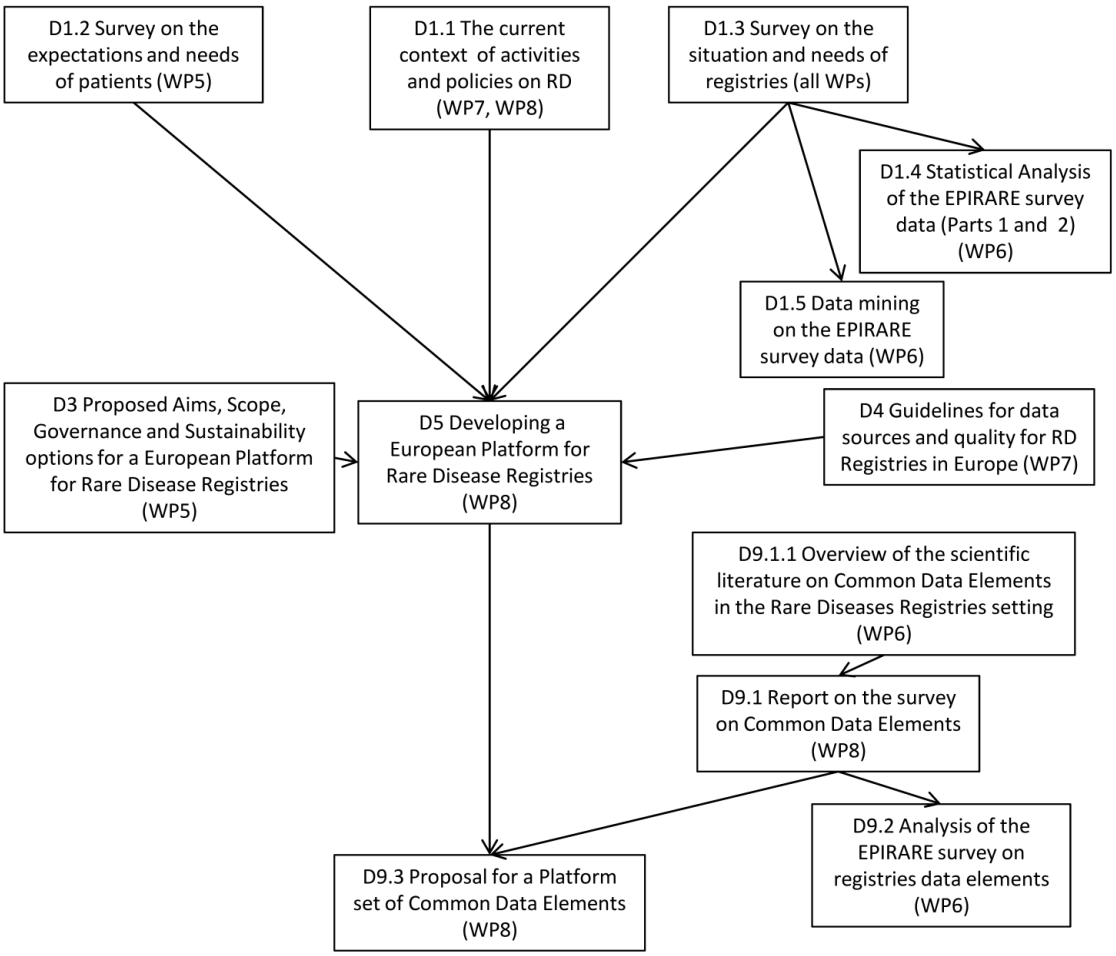
Epidemiological and public health considerations for the EPIRARE briefing document on RD and data protection.

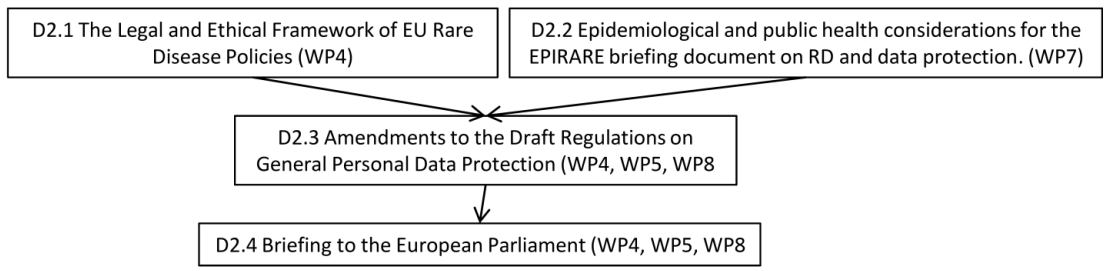
Manuel Posada

Institute of Rare Diseases Research (Instituto de Investigación de Enfermedades Raras - IIER), Instituto de Salud Carlos III (Instituto de Salud Carlos III - ISCIII); SpainRDR and Consortium for Biomedical Research in Rare Diseases (Centro de Investigación Biomédica en Red de Enfermedades Raras -CIBERER)

Contents

Overview of the documents produced by EPIRARE





Disclaimer

The contents of this document is in the sole responsibility of the Authors; The Executive Agency for Health and Consumers is not responsible for any use that may be made of the information contained herein.

Epidemiological and public health considerations for the briefing document on RD and data protection.

By Manuel Posada
June, 30th, 2012

Data sharing and research

One of the major achievements during the last 20 years has been the facility for the exchanging and communicating knowledge among people worldwide. This new possibility clearly opens new ways of communication and data sharing for basic scientists and clinical and epidemiologist researchers. However, this progress also opens the possibility of new personal risks due to confidentiality violation of the personal information. Many national and international committees and societies have expressed their concerns about the risk of developing strict rules and laws for data protection that do not allow to share scientific information, particularly those provide for epidemiological observational studies. In this sense, the International Ethics Committee of the Human Genome Organization (HUGO) stated in 2002 that human genomic databases should be considered as global public goods [1]. In this statement, global public goods were defined as goods 'whose scope extends worldwide, are enjoyable by all with no groups excluded, and when consumed by one individual, are not depleted for others' [2]. After this statement, several others such as the Fort Lauderdale rules of 2003 [3], the 2008 International Summit on Proteomics Data Release and Sharing Policy in Amsterdam [4] and the Toronto International Data Release Workshop of 2009 [5] have highlighted the importance of data sharing for the translational research at the global level. More recently, 17 major health funding agencies launched a joint statement about data sharing and public health [6]. They pointed out that in some research fields, data sharing has been well established and has accelerated the progress of research and its application for the public. The main three benefits that can be obtained from an appropriated data sharing among researchers are: faster progress in improving health, better value for money and higher quality science. , This statement also point out three main principles were based on equity, ethic and efficiency [7]. On the other hand, the international Code of Conduct elaborated by three international societies established the following principles for data sharing: Quality, Accessibility, Responsibility, Security, Transparency, Accountability and Integrity [8]

New European legislation on data protection

In January 2012, the EU published a draft Data Protection Regulation and Draft Data Protection Directive, as the first step towards legislative change [9]. Both the Directive and the Regulation offer new definitions for personal data which, in effect, widen the scope of the legislation. Increased emphasis is placed on the rights of the individual data subject to be fully informed and to understand the full extent of how their personal data is used. The requirements for consent are more explicit and robust. The 'genetic data' is

defined for the first time, as all data 'concerning the characteristics of an individual which are inherited or acquired during early prenatal development'. These increased rights afforded to the data subject are balanced by strengthened exemptions for data processing for health, public health (Article 81) and for 'historical, statistical and scientific research' (Article 83). These offer a more robust framework for those providing healthcare and medical research, and provide increased clarity about what the law allows [10].

However, some concerns have arisen about these two regulatory documents among public health researchers on Rare Diseases as well as patient organizations. In fact, the major risk of these type of regulations is that they try to protect personal data and individual data but at the same time they constrain very much the possibility of data sharing for research purposes. Specifically, rare diseases registries and biobanking activities are under much more stress than other research actions but they are fundamental for the development of the rare diseases knowledge.

On the occasion of the XIX IEA congress in Edinburgh in August 2011, a workshop discussed the current situation regarding data protection regulations and practices in the EU. It is argued that the revised version should take explicitly and adequately into account the special requirements for personal data collection, storage and use needed in epidemiological research activities such as health services auditing, studies involving disease registries and investigations of public health emergencies [11]. At the workshop it was unanimously recognized that it is necessary that the new proposal also acknowledges that epidemiological research is aimed at improving the health of populations and that not making use of available data would constitute a serious ethical problem, as stated in the "open letter of Nordic Countries [12].

Epidemiology and public health research

The main aim of epidemiological research is not to provide aggregated and descriptive data but to identify etiologic and risk factors for providing to the population the best prevention and new clues for new treatments. Therefore, epidemiological methods combine descriptive data just for generating hypothesis and analytical study designs for testing hypothesis. In all of these cases the fundamental principle of the epidemiology is the population because it is the focus on futures interventions, but the analysis is based on individual subjects because of things usually happens at individual level. Causality analysis takes into consideration whatever can happens one by one person due to genetic and familiar backgrounds, pregnancy and birth date features, environment, neuroconductual development, education, medications received and life styles, among some others. Indeed, social factors are also important and they intervene at population level but the responses to those factors have also an individual component.
Modern medicine is facing a new concept of disease. A disease is the combination of genetic background, environment and timing. In the future, we hope that a better understanding of how these features combine into patterns will generate new disease classifications, supporting greater specificity in health management techniques. Today, this trend toward greater specificity in health management based on detailed personal characteristics is commonly known as personalized health.

Rare Diseases research specificities

Rare diseases have been defined using the prevalence - an epidemiology estimator - at the European level. Some European citizens have different health and social rights just because they are a minority due to the diseases by which they are affected. Conscious of that, the EC has developed important regulations and recommendations to try to reduce differences to the health access services, reimbursement of health expenditures – including orphan drugs – and promoting research in this field. One of the main topic always mentioned in the rare diseases environments is related with the difficulty that researchers have to collect a sufficient number of cases for their investigations. Registries [13] and biobanking activities are very well recognized as important research infrastructures for the improvement of the rare diseases knowledge. Geographical analysis is also other important tool for detecting clusters and/or genetic founders.

Registries are in charge of collecting cases associated with identification and clinical data. Identification is absolutely needed because they are few cases affected by the same disease. They are undertaken to some delay diagnosis and during that period patients and families visit several centers searching for some solution. It is not uncommon that the same patient be included in several clinical and labs units of different centers belonging different regions within a country, and even in other countries. This is why, for registering and also for other research purposes it is imperative to have enough personal identification data in order to eliminate duplicates. For health and social planning purposes it is also important do not have duplicates in the estimation of the number of new cases (incidence), the number of cases alive at a point level (point prevalence) and the mortality risks of people affected by some specific RD.

Mortality information is one of the main concerns among data protection developers because of these two arguments,

i) Access to death certificate is not acceptable
ii) People dead should be deleted of a registry.

Regarding the first question, it is very important for a registry and also for cohort study designs to know if a patient is still alive. Surveillance is fundamental for the assessment of drugs and interventions (i.e.: transplants, major surgeries, population based screening etc) and quality methods assuring this knowledge needs for identification data.

However, modifications of certain aspects of death certificate registration and rules of data-protection are perhaps required to make international monitoring of place of death more feasible and accurate [14].

The second question is also a dilemma between those in favor of a strictly data protection rules and epidemiologists. It is not rare that researchers need to check back and review the data when a new hypothesis or question arises. If people already dead have been deleted of the registry, comparison between people still alive and people dead for searching prognosis factors would not be possible and several opportunities for new findings would have been lost.

Registries are important tools for collecting cases for new experiments (randomized clinical trials) and observational studies (including genetic studies). If patients included in a RD registry are anonymized, they would never have the possibility of being included in some of those studies. They never could give their consent for their participation because researchers did not know who is who.

Finally, the scarcity of cases for some RD claims for a worldwide cooperation, which is the major area of interest even from some European General Directorates which have signed international consortiums and approved projects addressed to these two topics: Registries and biobanks. The involvement of people from many countries and expertise are often needed to get the best possible results. The revision of the Directive will probably also involve new rules for moving data from EU to countries outside EU and this could be a problem if the epidemiological research is not considered. These revisions may simplify existing rules but could also make them more complicated [15].

Biobanking activities share some similarities with registries. They are in charge of collecting quality biological samples with clinical data associated. From the point of view of the data provided by donors, a sample is more or less equal to some other type of data to be included in a registry. However, the biological sample adds some extra possibilities and indeed creates important concerns regarding with data protection. The final destination of a biological sample is not to sleep in some high technological settlement for years but to be used by some researcher in a specific study. A biological sample contains important information, even personal genetic identification and background, though some extraction process may be needed to obtain that information. Regularly, biological samples are stored in their specific containers and freezers (depending of the type of sample) for years but marked with some type of barcodes. The information provided by this type of code does not contain personal data, but just information related to the type of sample, provider center, and date of extraction among, some other general information.. In other words, a biological sample does not contain readable personal data. However, the sample could be processed, if some analysis is done (particularly in DNA samples). Nevertheless, some genetic personal results would need to be compared with some population databank looking for some unique identification which is not already available for general population.
Geographical analysis has been also considered as an analytical tool where aggregation is commonly used. However, particularly for local phenomena that cross administrative boundaries, aggregation obscures spatial details needed for in-depth geographic analyses. Cluster analysis for rare cancers and congenital malformations need to use personal data for the identification and to study some susceptibility and behaviors attitudes. After collecting data and developing the corresponding statistical analysis, personal data is not need but should be saved for later checking [16].

References

1. Human Genome Organisation (HUGO), Ethics Committee: Statement on human genomic databases, December 2002. J Int Bioethique 2003, 14:207-210.

2. Human Genome Organisation (HUGO): Principles Agreed at the First International Strategy Meeting on Human Genome Sequencing: 25-28 February 1996; Bermuda. HUGO; 1996. [http://www.gene.ucl.ac.uk/hugo/bermuda.htm]

3. Sharing Data from Large-scale Biological Research Projects: A System of Tripartite Responsibility. Report of a meeting organized by the Wellcome Trust and held on 14-15 January 2003 at Fort Lauderdale, USA [http://www.sanger.ac.uk/datasharing/docs/fortlauderdalereport.pdf]

4. Rodriguez H, Snyder M, Uhlén M, Andrews P, Beavis R, Borchers C, Chalkley RJ, Cho SY, Cottingham K, Dunn M, Dylag T, Edgar R, Hare P, Heck AJ, Hirsch RF, Kennedy K, Kolar P, Kraus HJ, Mallick P, Nesvizhskii A, Ping P, Pontén F, Yang L, Yates JR, Stein SE, Hermjakob H, Kinsinger CR, Apweiler R: Recommendations from the 2008 International Summit on Proteomics Data Release and Sharing Policy: The Amsterdam principles. J Proteome Res 2009, 8:3689-3692.

5. Birney E, Hudson TJ, Green ED, Gunter C, Eddy S, Rogers J, Harris JR, Ehrlich SD, Apweiler R, Austin CP, Berglund L, Bobrow M, Bountra C, Brookes AJ, Cambon-Thomsen A, Carter NP, Chisholm RL, Contreras JL, Cooke RM, Crosby WL, Dewar K, Durbin R, Dyke SO, Ecker JR, El Emam K, Feuk L, Gabriel SB, Gallacher J, Gelbart WM, Granell A, et al.: Prepublication data sharing. Nature 2009, 461:168-170.

6. Sharing research data to improve public health: full joint statement by funders of health research [http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Public-health-and-epidemiology/WTDV030690.htm]

7. Mark Walport, Paul Brest. Sharing research data to improve public health. Lancet, 2007

8. Bartha Maria Knoppers, Jennifer R Harris, Anne Marie Tassé, Isabelle Budin-Ljøsne, Jane Kaye, Mylène Deschênes and Ma'n H Zawati. Towards a data sharing Code of Conduct for international genomic research. Genome Medicine 2011, 3:46

9. European Commission. A comprehensive approach on personal data protection in the European Union. COM. 609 final. Brussels: European Commission; 2010. http://ec.europa.eu/dataprotectionofficer/introduction_en.htm

10. Alison Hall. PHG Foundation. Proposed EU Data Protection reform and public health genomics, March 2012. http://www.phgfoundation.org/news/11462/

11. R. Saracci, J. Olsen, A. Seniori-Costantini, R. West. Epidemiology and the planned new Data Protection Directive of the European Union: A symposium report. Public Health. 126 (2012):253-255

12. Open letter to European Decision Makers. THE PLANNED CHANGES OF THE EUROPEAN DATA PROTECTION DIRECTIVE MAY POSE A THREAT TO IMPORTANT HEALTH RESEARCH. July 13, 2011. The 5th Nordic Meeting in Epidemiology and Register-based Health Research was held in Reykjavík, Iceland 14-16 June 2011. Available at "www3.svls.se/sektioner/svep/svepet/SVEPET_Nr2011-3.pdf"

13., Kastrup M, Mortensen PB. The Danish psychiatric register as a tool in epidemiology. Acta Psychiatr Scand Suppl. 1993;370:27-32.

14. Joachim Cohen, Johan Bilsen, Guido Miccinesi, Rurik Löfmark, Julia Addington-Hall, Stein Kaasa, Michael Norup, Gerrit van der Wal and Luc Deliens. Using death certificate data to study place of death in 9 European countries: opportunities and weaknesses. BMC Public Health 2007, 7:283

15. Jørn Olsen,Cesar Victora, Neil Pearce, Shah Ibrahim, Rodolfo Saracci. The IEA Rapid Response Committee. A new EU Directive for Data Protection in Europe. December 1, 2010. http://www.ieaweb.org/index.php?option=com_content&view=article&id=69:a-new-eu-directive-for-data-protection-in-europe&catid=22:rapid-response&Itemid=54

16. Kristen H. Hampton, Molly K. Fitch, William B. Allshouse, Irene A. Doherty, Dionne C. Gesink, Peter A. Leone, Marc L. Serre, and William C. Miller. Mapping Health Data: Improved Privacy Protection With Donut Method Geomasking. Am J Epidemiol 2010;172:1062–1069