

NAGEN-data: Analysis of the legal framework applicable to Genomic data discovery for secondary use of personal data.

Armendia-Girón D, Jorqui-Azofra M, Luquin-Bergareche L, Alvira-Iraizoz F, Almanza-Goicoechea A, Alonso-Sanchez A, Egusquiza-Balmaseda MA†, García-Solaesa V†

* All authors declare no competing interest † These authors jointly supervised this work

Background: NAGEN program has collated more than 3000 genomes in consecutive studies of complex hereditary diseases. In this context, NAGEN-data aims to create an ecosystem of genomic and phenoclinical data allowing access for research purposes via data discovery tools like Beacon v2. From the ethical-legal perspective, the protection of personal data needs to be considered, not only due to the evident importance of the regulatory framework applicable to genetic data, but also due to the future implementation of the European Health Data Space (EHDS) by the EU. Therefore, our objective was to study the main issues that affect, from the perspective of regulatory compliance and the guarantee of personal data protection, the deployment of an ecosystem, such as the one described, by default and by design. **Methods:** We thoroughly revised the legal framework applicable to the NAGEN-data ecosystem governed by the General Data Protection Regulation (GDPR), the Data protection Act (LOPDGDD), the Biomedical Research Act as well as the future EHDS. **Results:** Despite the lack of entry into force of EHDS regulation, the current regime allows the creation of the envisioned ecosystem. The Spanish contribution to the regime requires the adoption of concrete technical and organizational measures for the pseudonymization and minimization of data and the implementation of a Data Protection Impact Assessment (DPIA). In this sense, we considered Beacon an appropriate tool to guarantee compliance with pseudonymization and minimization of data. **Conclusions:** Specific data access tools such as Beacon must be developed and implemented to guarantee regulatory compliance provided by legislators, which must be considered from the design of the genomic data-sharing ecosystem.

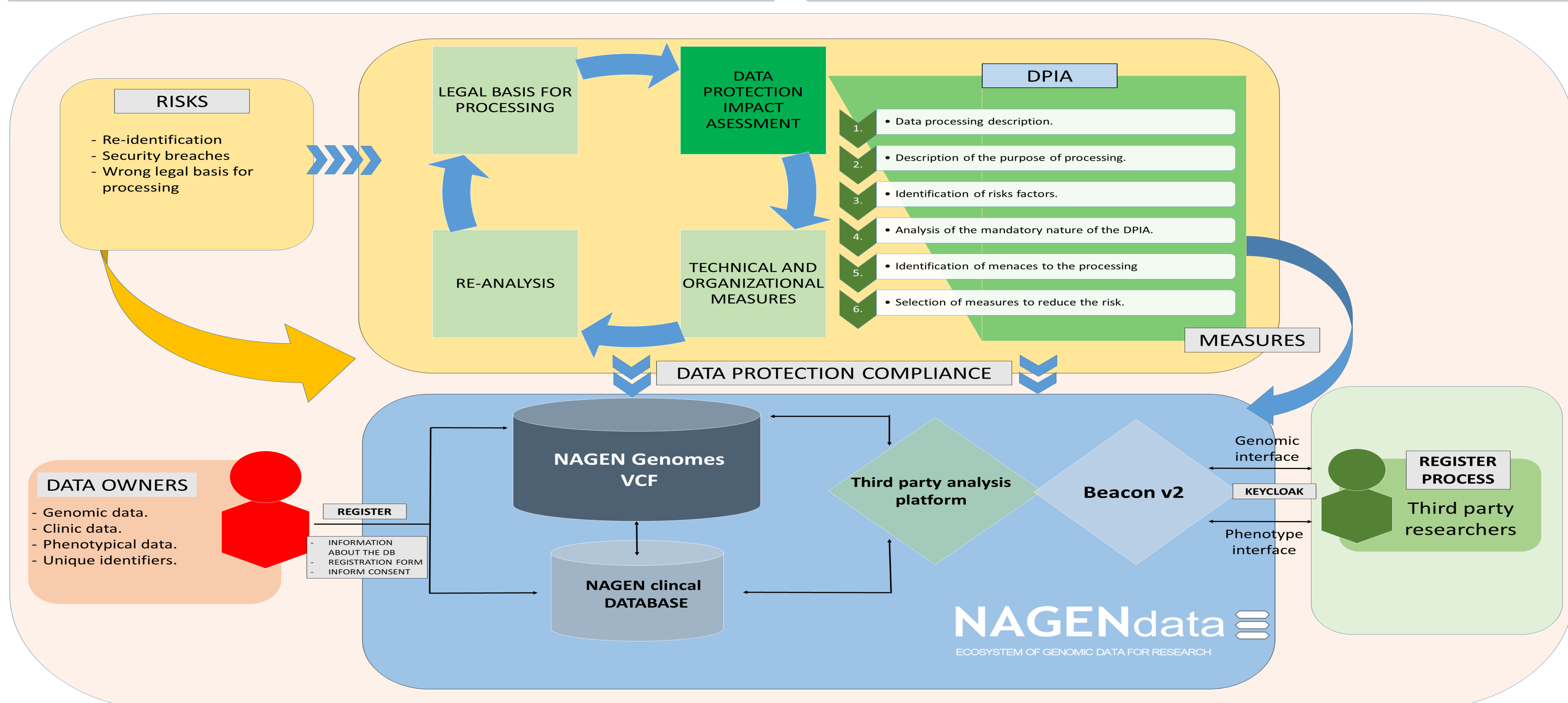
Keywords: Genomic data infrastructure, Data discovery, Legal framework, Data Protection impact assessment, Beacon.

Introduction: NAGEN-data faces the importance of sharing and reusing health data in a research context by creating an equally accessible infrastructure for other biomedical researches. While the aim of the ecosystem lines up with universal biomedical principles, it is unclear how it behaves in the current legal framework. Currently, and taking into account the future entry into force of the European Health Data Space, both the European and National legislators provides a compendium of four main legal Acts: the General Data Protection Regulation, the Spanish Data protection Act (Ley Orgánica de Protección de Datos Personales y Garantía de los Derechos Digitales, LOPDGDD), the Spanish Biomedical Research Act and last but not least the European Health Data Space Regulation, recently approved by the European Parliament and the Council. We made an in-depth, ground-up legal analysis of the current legal framework regulating data processing in biomedical research in Spain.

Results: Through the in-depth legal analysis conducted for NAGEN-data, a number of “high impact” areas have been identified where thoroughness is required to ensure regulatory compliance:

- 1. Identification of the legal basis for the processing:** Due to the presence of “special categories of personal data” (Article 9.1 GDPR) the processing of these data is generally prohibited. However, the legislator provides a series of exceptions to this prohibition, including the processing of personal data for scientific research or historical purposes. This circumstance is complemented by the specific regulation of the Spanish Data Protection Act which adds additional conditions in Additional Provision 17, which must be taken into account.
- 2. Adoption of technical and organizational measures to guarantee regulatory compliance:** Due to the nature of the data and the legal basis for its processing, the legislator requires those who process personal data for research purposes to be especially careful to safeguard the rights of the data “owners.” Therefore, throughout the legislation, it reiterates the need to adopt measures to guarantee, for example, the pseudonymization of personal data. Taking these factors into account, the team has implemented, among others, Beacon, a discovery tool approved by GA4GH, and KeyClock, a software for access control, as fundamental pillars to maintain governance or measures related to it.
- 3. Conducting a Data Protection Impact Assessment:** Due to the combination of regulations applicable to the ecosystem, a DPIA is necessary. This legal tool aims to analyse the data processing carried out by NAGEN-data throughout its entire lifecycle, to identify potential risks and threats and to adopt control measures aimed at minimizing their appearing chances. The measures adopted in this context have been aimed at minimizing the risks of re-identification, as well as establishing access control measures, among other.

Conclusions: Due to the regulatory context that applies to personal data processing in biomedical research, like in the NAGEN-data ecosystem, it is necessary to clearly justify the basis of data processing. In this framework, the legal basis of scientific research implies to guarantee regulatory compliance with the adoption of technical and organizational measures to fulfil the content of the data protection principles and an analysis of the risks and threats of the data processing in order to protect the rights of the data “owners”. In our case, Beacon and Keycloak acts as perfect tools to establish a standard in the technical side of the ecosystem, while the DPIA provides us with organizational and design responses to the analysis of the risks and threats to the rights of the data “owners”.



Contact information:
Daniel Armendia Girón
danielarmendiagiron@gmail.com
Unit of Genomics Medicine - Navarrabiomed

NAGENdata 
ECOSYSTEM OF GENOMIC DATA FOR RESEARCH