Public and patient involvement in health data governance (DATAGov): protocol of a people-centred, mixed-methods study on data use and sharing for rare diseases care and research

Cláudio de Freitas,1,2 Mariana Amorim,1 Helena Machado,3 Elisa Leão Teles,4 Maria João Baptista,5,6 Alicia Renedo,7 Veerle Provoost,8 Susana Silva1,9

ABSTRACT

Introduction International policy imperatives for the public and patient involvement in the governance of health data coexist with conflicting cross-border policies on data sharing. This can challenge the planning and implementation of participatory data governance in healthcare services locally. Engaging with local stakeholders and understanding how their needs, values and preferences for governing health data can be articulated with policies made at the supranational level is crucial. This paper describes a protocol for a project that aims to coproduce a people-centred model for involving patients and the public in decision-making processes about the use and sharing of health data for rare diseases care and research.

Methods and analysis This multidisciplinary project draws on an explanatory sequential mixed-methods study. A hospital-based survey with patients, informal carers, health professionals and technical staff recruited at two reference centres for rare diseases in Portugal will be conducted first. The qualitative study will follow consisting of semi-structured interviews and scenario-based workshops with a subsample of the participant groups recruited at baseline. Quantitative data will be analysed using descriptive and inferential statistics. Inductive and deductive approaches will be combined to analyse the qualitative interviews. Data from scenario-based workshops will be iteratively compared using the constant comparison method to identify cross-cutting themes and categories.

Ethics and dissemination The Ethics Committee for Health from the University Hospital Centre São João/ Faculty of Medicine of University of Porto approved the study protocol (Ref. 99/19). Research findings will be disseminated at academic conferences and science promotion events, and through public meetings involving patient representatives, practitioners, policy-makers and students, a project website and peer-reviewed journal publications.

INTRODUCTION

How to govern health data has become a question of unprecedented relevance in the golden age of data sharing.1 Large-scale health data repositories aggregating multiple datasets hold enormous predictive and transformation potential to improve population health and well-being.2 This is especially promising in the field of rare diseases, where transnational data sharing is likely to intensify as a result of increased sponsorship by the European Commission (EC).3 Yet how health data are shared and used also carries risks and causes public concern.4 International policy agencies recommend public participation in
data governance as a means to improve safeguards, reduce risks and increase the benefits arising from health data processing.\(^5\)\(^6\) But such recommendations exist alongside conflicting cross-border policies on data sharing\(^2\)\(^3\) that can thwart its implementation. Engaging with local stakeholders and understanding how their needs, values and preferences can be articulated with policies made at the supranational level is therefore a pressing concern.

Processing of large sets of health data can help to prevent, diagnose and cure hundreds of diseases.\(^9\)\(^10\) It can also improve clinical practice and quality of care, reduce public health threats and advance health policy goals.\(^5\)\(^9\)\(^11\)\(^12\) However, information security and privacy have been challenged by data breaches and mismanagement that can lead to participants’ undue identification.\(^13\)\(^14\) Unforeseen access to health data by private companies also raises concerns with confidentiality and inappropriate data use, including results commercialisation and undisclosed surveillance.\(^4\)\(^15\)\(^16\) These risks can erode public trust in the organisations that handle their data, reducing potential for innovation.\(^4\)\(^17\) To address these issues, several international legal instruments and recommendations have been produced.\(^5\)\(^6\)\(^18\) They emphasise the importance of involving the public in the development of a shared policy framework to govern the protection of privacy, data security and transnational data flows.\(^5\)\(^6\)

Rare diseases offer a paradigmatic example of the opportunities and challenges involved in the governance of health data. Their rarity, diversity and complexity demand collaboration and sharing of patient data between experts internationally. Data-sharing partnerships are key to overcome difficulties of conducting research in this field: case numbers are low and there is insufficient availability of technological tools needed to conduct genetic testing and genome sequencing.\(^19\)\(^20\) Yet this type of collaboration is difficult as data on rare diseases is sparse, dispersed and unsystematically collected and stored.\(^21\) To tackle this problem, the EC has recently funded the European Reference Networks (ERN) to set up rare disease patient registries.\(^3\) ERNs are expected to develop interoperable registries and to build synergetic partnerships for knowledge and data exchange.\(^22\) Realising these goals will prompt the flow of large amounts of data and technologies between member states.

Cross-border health data sharing raises additional ethical, legal and social challenges. These include unauthorised data reuse and unwanted return of incidental findings from research,\(^4\)\(^23\) which compromise individuals’ autonomy and entitlement to an open future (eg, anticipating the diagnosis of an untreatable adult-onset disorder through return of unsolicited genetic testing results).\(^24\) The EU General Data Protection Regulation, enforced since May 2018, aims to address some of these challenges, not least through the rights to be forgotten and to data portability.\(^18\) It does so on the footprints of two sets of policies on research and innovation. One of those policies promotes data sharing and harmonisation of international rules toward a paradigm of ‘open science’.\(^25\) The other one is underlined by ethical concerns with autonomy and privacy protection and asserts the need to control data flows.\(^26\) These data-sharing policies host mutually incompatible rules that can give rise to tensions when implemented:\(^27\) healthcare providers, for example, may struggle to combine the protection of patients’ right to privacy while seeking to abide by the principles of open-access.

**Participatory health data governance**

Increasing support for participatory health data governance, whereby patients and other members of the public are involved in consultations and decision making about health data, is premised on the idea that this process can increase accountability in data processing, reduce unnecessary harms and enhance the fair distribution of the benefits of data use.\(^5\)\(^6\)\(^8\)\(^28\) Participatory data governance is also encouraged as a way of promoting dialogue about ethical, legal and social issues and imbuing research policy with socioethical sensitivity.\(^29\)–\(^32\) Collaboration between lay and research experts can infuse research with social values and facilitate the codesign of innovative solutions for complex problems.\(^30\)\(^31\) However, those collaborative relationships are not always easy to develop.\(^33\) Careful attention is needed to create participatory spaces that are inclusive and harness everyone’s potential for contribution.\(^34\)–\(^36\) Competing interests, and an unbalanced distribution of the resources needed for meaningful participation, can cause disadvantaged groups to be excluded from decision making, which in turn risks to reproduce ethnic, age, gender and socioeconomic inequalities or to even generate new ones (eg, digital exclusion).\(^34\)\(^37\)–\(^40\) There is thus an unequivocal need to assess the potential benefits and risks of public and patient involvement in health data governance and to develop a people-centred, bottom-up approach to participation that responds to local stakeholders’ interests and needs. This project aims to reach this goal by collaborating with stakeholders engaged in care and research for rare diseases at an academic hospital centre in Portugal. Its specific aims are to:

1. Assess the needs and preferences of rare diseases patients, their informal carers, health professionals and technical staff concerning decision making about the use and sharing of health data for rare diseases care and research.
2. Understand those stakeholders’ expectations of and perspectives about public and patient involvement in health data decision-making processes, including their views on the ethical, legal and social implications of participation.
3. Coproduce a people-centred model for public and patient involvement in health data governance.

DATAGov is led by a multidisciplinary team working at the crossroads of sociology, public health and clinical medicine and with long-term experience in collaborating with rare diseases community organisations and networks. The project will contribute to advance international...
METHODS AND ANALYSIS

Study design and setting

This project adopts an explanatory sequential mixed-methods design, collecting and analysing quantitative and qualitative data.41 The quantitative phase is first in the sequence and draws on a hospital-based survey with rare disease patients, informal carers, health professionals and technical staff recruited at the University Hospital Centre São João (UHCSJ), in Porto, Portugal. The qualitative phase relies on semistructured interviews and scenario-based workshops with a subsample of the participants recruited at baseline. Survey and interview data will be collected to examine stakeholders’ needs, preferences, expectations and perspectives of public and patient involvement in health data decision making. These findings will be subsequently integrated to develop a set of scenarios for participatory health data governance, which will be presented at the scenario-based workshops. Representatives of all stakeholder groups will be invited to discuss the scenarios and to select a preferred model for involving patients and other members of the public in the governance of health data (see table 1). Integration of quantitative and qualitative findings will provide an understanding of the ethical, legal and social implications of participatory data governance.

Data will be collected in two reference centres for rare diseases at the UHCSJ. These research settings were purposefully sampled because: (1) they were among the first reference centres to be created in Portugal in 2016 and are acknowledged for their expertise and clinical practice; (2) they are in charge of overseeing patients from the entire Northern Health Region of Portugal, which allows for larger study samples; (3) one is affiliated with an ERN and the other is awaiting a final decision on its application for membership and (4) they are involved in national and international research projects, which presents an excellent case to explore first-hand experiences of decision making regarding patient data use and sharing.

Study procedures and analysis

Hospital-based survey

Rare disease patients, informal carers, health professionals and technical staff will be invited to participate in a survey conducted at the UHCSJ to assess their preferences and needs regarding decision making concerned with health data use and sharing.

During a 6-month period, patients aged 12 and above who attend a consultation in the reference centres at the UHCSJ, who can read and write in the Portuguese language and without cognitive impairment will be consecutively recruited to participate in the study, together with their informal carers. Medicine and nursing unit staff of the reference centres outpatient clinics will distribute study information leaflets to potential participants, or their guardians, after patients’ biometric screening appointments or medical consultations. Two researchers will be at the hospital site to provide additional information and to invite patients and informal carers to independently fill in a self-report structured questionnaire, after obtaining informed consent. Underage participants will be invited to complete the questionnaire individually, after giving verbal consent and their carers signing the written consent form. Adult survey participants will be asked for consent to be contacted and invited to a semistructured interview approximately 4 months later.

Directors and managers of relevant units and departments at the UHCSJ will be provided with recruitment emails, along with study information sheets, to invite health professionals from the reference centres’ multidisciplinary teams (ie, doctors, nurses and allied health professionals) and patient representatives to participate in the study procedure.

Data collection will be conducted in two reference centres for rare diseases at the UHCSJ. These research settings were purposefully sampled because: (1) they were among the first reference centres to be created in Portugal in 2016 and are acknowledged for their expertise and clinical practice; (2) they are in charge of overseeing patients

<table>
<thead>
<tr>
<th>Table 1 Mixed-methods research objectives, procedures and analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Objective</strong></td>
</tr>
<tr>
<td>To assess stakeholders’ needs and preferences concerning decision-making about health data use and sharing</td>
</tr>
<tr>
<td>To understand stakeholders’ expectations and perspectives about public and patient involvement in health data decision-making processes</td>
</tr>
<tr>
<td>To coproduce a model for public and patient involvement in health data governance</td>
</tr>
</tbody>
</table>
Categorical variables will be assessed using the appropriate tests, and differences in frequencies and proportions of categorical variables will be evaluated with the Student's t-test or Mann-Whitney test, as appropriate. Analytic approaches will be performed first using open, selective and focused analysis to tackle the variation in perspectives and experiences and allow theoretical saturation to be reached. Interviews will be held at a time and place of participants’ preference (eg, university department, work place, home). Participant information sheets and consent forms will be provided alongside with a request to audiorecord the interviews.

Questions for the interview guide are exploratory and informed by existing literature. They address four main topics: (1) personal experience with involvement in patient organisations and/or illness; (2) facilitators of and barriers to health data sharing; (3) views about individual participation in decision-making concerning health data use and sharing and, (4) perspectives about public and patient involvement in health data governance. Interviews are expected to last between 30 min and 1 hour. We estimate to conduct interviews with 30 patients, 30 informal carers, 15 health professionals and 10 technical staff.

Data analysis will proceed hand in hand with data collection for timely identification of informational redundancy and subsequent discontinuation of sampling. Interview transcripts will be analysed iteratively by two researchers with the support of NVivo V.12. Inductive analysis will be performed first using open, selective and axial coding to facilitate the emergence of categories and core themes. Deductive analysis will follow drawing on relevant theoretical frameworks, namely the Modified Participation Chain Model (MPCM). The MPCM asserts that public involvement in health decision making is influenced by a set of mutually constitutive factors that evolve through participation when the right institutional dynamics are in place. These factors include individual and collective motivations, mobilisation efforts, resources and the dynamics of participation. They influence not only people’s willingness to get involved but also their capacity and resolve to participate. These acumens will be used to load inductive themes with theoretical sensitivity. Triangulation of interview data sources and of data analysts will be employed to ensure the rigour and quality of the findings.

Stakeholders’ views will provide in depth insight in:

- How they may be driven to or detracted from data use and sharing for care and research and whether that is explained by perceived benefits and risks?
- Their stand concerning the involvement of different publics in collective decision-making about health data.
- Their views about the ethical, legal and social implications of participatory health data governance and proposed solutions to overcome potential challenges.

Qualitative interviews

To deepen our understanding of stakeholders’ expectations and perspectives of public and patient involvement in health data decision making, semi-structured interviews will be conducted with a subsample of the adult participants who completed the questionnaire. Approximately 4 months after participation in the survey, patients and informal carers who consented to being contacted will be invited by telephone or email to take part in an individual qualitative interview. Health professionals and technical staff will be invited to participate through email. Heterogeneity sampling will be employed to obtain maximum variation in perspectives and experiences and allow theoretical saturation to be reached. Interviews will be held at a time and place of participants’ preference (eg, university department, work place, home). Participant information sheets and consent forms will be provided alongside with a request to audiorecord the interviews.

Questions for the interview guide are exploratory and informed by existing literature. They address four main topics: (1) personal experience with involvement in patient organisations and/or illness; (2) facilitators of and barriers to health data sharing; (3) views about individual participation in decision-making concerning health data use and sharing and, (4) perspectives about public and patient involvement in health data governance. Interviews are expected to last between 30 min and 1 hour. We estimate to conduct interviews with 30 patients, 30 informal carers, 15 health professionals and 10 technical staff.

Data analysis will proceed hand in hand with data collection for timely identification of informational redundancy and subsequent discontinuation of sampling. Interview transcripts will be analysed iteratively by two researchers with the support of NVivo V.12. Inductive analysis will be performed first using open, selective and axial coding to facilitate the emergence of categories and core themes. Deductive analysis will follow drawing on relevant theoretical frameworks, namely the Modified Participation Chain Model (MPCM). The MPCM asserts that public involvement in health decision making is influenced by a set of mutually constitutive factors that evolve through participation when the right institutional dynamics are in place. These factors include individual and collective motivations, mobilisation efforts, resources and the dynamics of participation. They influence not only people’s willingness to get involved but also their capacity and resolve to participate. These acumens will be used to load inductive themes with theoretical sensitivity. Triangulation of interview data sources and of data analysts will be employed to ensure the rigour and quality of the findings.

Stakeholders’ views will provide in depth insight in:

- How they may be driven to or detracted from data use and sharing for care and research and whether that is explained by perceived benefits and risks?
- Their stand concerning the involvement of different publics in collective decision-making about health data.
- Their views about the ethical, legal and social implications of participatory health data governance and proposed solutions to overcome potential challenges.

Scenario-based workshops

Scenario-based workshops will be held to coproduce a people-centred model for public and patient involvement
in health data governance. This foresight method was chosen for its potential to facilitate the engagement of patients and the public in the development of health policy at local and national levels, and it will be used to incorporate participants’ collective expertise into a model for participatory data governance.

Survey and interview data will be integrated and combined with extant literature to inform the design of a set of scenarios for involving different stakeholders in collective decision-making processes concerned with health data. The scenarios will be presented in workshops with representatives from rare diseases patients, informal carers, health professionals and technical staff to enable critical dialogue. The scenario-based workshops will be carried out in two rounds: (1) round 1 will entail four intra-group workshops conducted with each stakeholder group separately (8–12 participants). Participants will be asked to evaluate the relevance and feasibility of each scenario and to assess its potential ethical, legal and social implications; (2) round 2 will entail two mixed-group workshops including a similar number of representatives from each stakeholder group (n=3) in a total of 12 participants. Stakeholders’ feedback from round 1 will be used to make revisions to the first set of scenarios. The revised scenarios will be offered for discussion in round 2, in which participants will be asked to select a preferred model for public and patient involvement in health data governance.

Round 1 participants will be selected per stakeholder group and from the baseline sample through heterogeneity sampling, stratified by gender and educational level. They will be invited to participate by email or telephone, following their expressed preferences. A subsample of round 1 participants will be randomly selected to take part in round 2. The scenario-based workshops will be held in the venues considered most accessible by participants (eg, university department). They will be moderated by an experienced researcher supported by two observers who will carry out observation and note-taking. They are expected to last approximately 2 hours. Permission to audiorecord the workshops will be requested. Recordings will be transcribed verbatim and used for interpretational analysis with the assistance of NVivo V.12. Following open, axial and selective coding, data from each scenario-based workshop will be iteratively compared with data from other, same-round scenario-based workshops using the constant-comparison method to identify cross-cutting core themes and categories.

The scenario-based workshops will offer insights about:

- What motivates stakeholders to get involved in collective decision making about health data?
- Which resources are needed to promote meaningful participation by all stakeholders involved?
- How stakeholders can be best informed about and mobilised to undertake opportunities for involvement in health data governance?
- How collaborative participation dynamics may be achieved so as to accommodate stakeholders’ needs, values and preferences and inform a people-centred model for participatory data governance?

Public and patient involvement

The project’s research question was informed by a previous empirical study conducted by some of the authors in which rare diseases community experts expressed the need to play an active role in decisions about how their health data is used. Representatives of patients, informal carers, health professionals and researchers were involved in questionnaire and interview guide validation through pretesting and piloting. Their feedback was used to adapt its contents and to keep their length to a feasible minimum. In addition, the following strategies will be used to promote inclusivity in study participation: participants will be recruited by health professionals in a consecutive way, after their medical appointments; research team members will support participants during questionnaire completion, clarifying any arising doubts and interviews and scenario-based workshops will accommodate participants’ scheduling and place preferences (eg, weekends; close to work).

The project’s ultimate aim is to facilitate the co-production of a model for participatory health data governance. This will be achieved through involvement of a variety of stakeholders in critical dialogue about how to govern health data using scenario-based workshops. Participatory exercises that engage stakeholders with different backgrounds and interests offer an opportunity to expand knowledge and effect change that matters to increasingly diverse constituencies. However, they may also cause the powerful-powerless gap to become more salient if the voices of groups traditionally lacking in equitable social participation are silenced. To prevent the latter, patients and informal carers will be provided with a training in preparation for the scenario-based workshops. All participants will be informed of the purposes and dynamics (eg, moderators’ and participants’ roles; appropriate rules of conduct) of the scenario-based workshops prior to their start. Furthermore, intragroup scenario-based workshops will be held first, to ensure that the perspectives of all stakeholder groups are elicited and incorporated in the scenarios presented at the mixed-group workshops. All workshops will be facilitated by academic leads with experience in involving patients and the public in healthcare research.

Stakeholder representatives will also be involved in research findings dissemination through participation in public meetings and science promotion events. The project website will provide a contact platform that stakeholders and the general public can use to address questions and comments to the research team.

ETHICS AND DISSEMINATION

The study protocol received ethical approval from the Ethics Committee for Health from the UHCSJ/Faculty of Medicine of University of Porto (Ref. 99/19).
To ensure data confidentiality and protection, participants’ identifying information will only be available to four research team members, who will sign a non-disclosure agreement. Other researchers will be provided pseudonymised data. Questionnaires, interviews and scenario-based workshops will be conducted in a private place after obtaining consent. Personal contacts, consent forms, questionnaires, audio recordings and transcripts will be coded and kept separately in computer files protected by password and in locked cabinets. Audio and transcribed data will be securely stored for a period of 5 years and eliminated afterwards.

Survey and interview research findings will be reported first to stakeholders at the in-group scenario-based workshops to identify areas of convergence and tension and prepare for inter-group dialogue about participatory data governance. In addition, the findings will be disseminated to a broad audience through peer-reviewed journal publications, academic conferences, a project website and science promotion events involving patient representatives, practitioners, policy-makers and students.

The project was initially planned to last 36 months, starting in July 2018. However, an extension of 12 months was requested to compensate for fieldwork constraints caused by the COVID-19 pandemic. The project is thus expected to run from July 2018 to July 2022. Adjustments to data collection and dissemination, including the use of online tools (eg, online questionnaires, webinars), will be made if the measures undertaken locally to address the pandemic so require.

EXPECTED OUTCOMES AND RELEVANCE

International policy recommendations for public involvement in data governance exist amidst conflicting data-sharing policies. This can challenge the effective planning and implementation of participatory data governance in local healthcare services. DATAGov is set to coproduce a people-centred model for collective participation in decision-making about data use and sharing for rare diseases care and research. To that end, it will help build collective expertise by a wide range of stakeholders, including patients and lay members of the public, both to unravel the social, legal and ethical implications of participatory data governance and to devise bottom-up strategies to address emerging problems.

Project results will contribute to inform and improve the practice of public and patient involvement in health data governance, nationally and internationally, by expanding the evidence base on stakeholders’ stands, concerns and preferred roles regarding data decision-making processes. Undeniably, DATAGov is a timely and much needed response to calls for moving away from ‘hypothetical person-centred approaches that presume a universal “reasonable person” standard’ into empirically grounded accounts of data governance that acknowledge the heterogeneity of interests underlining the constitution of different publics. As such, it will help to reduce policy implementation gaps by facilitating the coproduction of a data governance model more attuned with citizens’ expectations, needs and preferences.
3 European Commission. Support for the setting-up of registries of patients affected by rare diseases available for all the ERNs, 2019.