Clinical Microbiology and Infection 31 (2025) 753–760

Contents lists available at ScienceDirect

Clinical Microbiology and Infection

journal homepage: www.clinicalmicrobiologyandinfection.com

Narrative review

How to: share and reuse data—challenges and solutions from predicting the impact of monoclonal antibodies & vaccines on antimicrobial resistance project

Mariana Guedes [1, 2, 3, *], Almudena de la Serna Bazan [1, 2, 3], Elena Rubio-Martín [1, 2, 3, 4], Lydia Barrera Pulido [1, 2, 3], Virginia Palomo [1, 2, 3, 4], Alen Piljić [5], Quentin J. Leclerc [6, 7, 8], Emmanuel Aris [9], Venanzio Vella [10], Asta Dambrauskienė [11, 12], Julie V. Robotham [13], Astrid Pérez [14], Nasreen Hassoun-Kheir [15, 16], Marlieke E.A. de Kraker [15, 16], Fabiana Arieti [17], Ruth Joanna Davis [17], Evelina Tacconelli [17], Elena Salamanca-Rivera [1, 2, 3, 4], Jesús Rodríguez-Baño [1, 2, 3, 4]

[1] *Infectious Diseases and Microbiology Division, Hospital Universitario Virgen Macarena, Seville, Spain*
[2] *Department of Medicine, University of Sevilla/Instituto de Biomedicina de Sevilla (IBiS)/Consejo Superior de Investigaciones Científicas (CSIC), Seville, Spain*
[3] *Instituto de Biomedicina de Sevilla (IBiS)/Consejo Superior de Investigaciones Científicas (CSIC), Seville, Spain*
[4] *Centro de Investigación Biomédica en Red Enfermedades Infecciosas (CIBERINFEC), Instituto de Salud Carlos III, Madrid, Spain*
[5] *Life Science Network, Heidelberg, Germany*
[6] *Institut Pasteur, Université Paris Cité, Epidemiology and Modelling of Bacterial Escape to Antimicrobials (EMEA), Paris, France*
[7] *Laboratoire Modélisation, Epidémiologie et Surveillance des Risques Sanitaires, Conservatoire National des Arts et Métiers, Paris, France*
[8] *Institut National de la Santé et de la Recherche Médicale (INSERM), Université Paris-Saclay, Université de Versailles St-Quentin-en-Yvelines, Team Echappement aux Anti-infectieux et Pharmacoépidémiologie U1018, Team Echappement aux Anti-infectieux et Pharmacoépidémiologie (CESP), Versailles, France*
[9] *GlaxoSmithKline (GSK), Wavre, Belgium*
[10] *GlaxoSmithKline (GSK), Siena, Italy*
[11] *Infection Control Service, Hospital of Lithuanian University of Health Sciences Kauno klinikos, Kaunas, Lithuania*
[12] *Department of Laboratory Medicine, University of Health Sciences, Kaunas, Lithuania*
[13] *Clinical and Public Health, United Kingdom (UK) Health Security Agency, London, United Kingdom*
[14] *Department of Infections Associated to Health Care, Spanish Centre for Microbiology (CNM)-Institute of Health Carlos III, Madrid, Spain*
[15] *Infection Control Program, Geneva University Hospitals and Faculty of Medicine, Geneva, Switzerland*
[16] *WHO Collaborating Centre, Geneva, Switzerland*
[17] *Division of Infectious Diseases, Department of Diagnostics and Public Health, University of Verona, Verona, Italy*

ARTICLE INFO

Article history:
Received 22 October 2024
Received in revised form
3 January 2025
Accepted 20 January 2025
Available online 25 January 2025

Editor: L. Leibovici

Keywords:
Anonymized data
Antimicrobial resistance
Data repository
Data reuse
Data sharing
FAIR

ABSTRACT

Background: Data sharing accelerates scientific progress and improves evidence quality. Even though journals and funding institutions require investigators to share data, only a small part of studies made their data publicly available upon publication. The procedures necessary to share retrospective data for reuse in secondary data analysis projects can be cumbersome.
Objectives: Predicting the Impact of Monoclonal Antibodies & Vaccines on Antimicrobial Resistance is a European research project that aims to develop mathematical models and an epidemiological repository to assess the impact of vaccines and monoclonal antibodies on antimicrobial resistance (AMR). To accomplish the project aim, Work Package 3 was responsible for gathering historical anonymized individual patient datasets.
Sources: Through a systematic search we have identified 108 eligible studies for data sharing; of which eight have completed all legal requirements and shared their datasets, with data from four infectious syndromes and seven resistant pathogens. The AMR data gathered in Predicting the Impact of Monoclonal Antibodies & Vaccines on Antimicrobial Resistance project are publicly available in European Clinical Research Alliance on Infectious Disease epidemiology network platform (https://epi-net.eu/primavera/about/anonymized-individual-patient-data/).

* Corresponding author: Mariana Guedes, Infectious Diseases and Microbiology Division, Hospital Universitario Virgen Macarena, Av. Dr. Fedriani, 3, 41009 Seville, Spain.
E-mail address: mariananpguedes@gmail.com (M. Guedes).

https://doi.org/10.1016/j.cmi.2025.01.024

Individual patient data
Secondary data analysis

Content: Challenges and possible solutions in data sharing activities were mapped and discussed: lack of researchers' interest in sharing data, cumbersome ethical and legal requirements, laborious data management procedures, specific requirements for public data access, insufficient training and funding.
Implications: We expect that experience gained in this project can be useful to improve data sharing; and that the datasets gathered can be used in future AMR projects. **Mariana Guedes, Clin Microbiol Infect 2025;31:753**

Background

Data sharing increases public trust in research by enhancing transparency and allowing independent researchers to replicate original analyses [1,2]. It also enables researchers to explore new hypotheses and gain new medical insights while reducing unnecessary duplication of studies [1–3] Over the last decade, journals and funding institutions have increasingly required investigators to share data [2,4]. The International Committee of Medical Journal Editors has proposed that authors share deidentified patient data underlying their results upon publication; and regulatory bodies, such as the European Medicines Agency, have published clinical data to prevent trials duplication, foster innovation and build trust in regulatory processes [4,5]. Additionally, findable, accessible, interoperable, reusable (FAIR) data principles were established [6].

In recent years, data repositories and portals have emerged. Some examples include Figshare platform, Dryad data repository, Yale University Open Data project and Vivli platform [2,7–9]. Other initiatives focused on specific diseases or developed by certain consortia, such as the Worldwide Antimalarial Resistance Network, National Heart, Lung and Blood Institute data repository, Virtual International Stroke Trials Archive or Alzheimer's Disease Neuroimaging Initiative facilitate sharing data in a collaborative environment [10–13].

The 2023 State of Open Data survey indicated that 37% of responders share their data upon publication, 18% by end of the research cycle, and 17% only upon request [14]. When compared with other research fields, medicine and social sciences were significantly more likely to share their data only upon request [14]. On the other hand, researchers are interested in re-using data, with data portals receiving a substantial number of data requests [11,15,16].

Aim of predicting the impact of monoclonal antibodies & vaccines on antimicrobial resistance project

Predicting the Impact of Monoclonal Antibodies & Vaccines on Antimicrobial Resistance (PrIMAVeRa; https://www.primavera-amr.eu/) is a European private-public partnership project funded by Innovative Medicines Initiative 2 Joint Undertaking that aims to develop mathematical models and an epidemiological repository to assess the impact of vaccines and monoclonal antibodies (mAbs) on antimicrobial resistance (AMR) [17]. PrIMAVeRa work package 3 (WP3) objectives included identify existing datasets with relevant individual patient data on infections caused by AMR bacteria; assess the accessibility of individual patient data and develop a data management and validation plan; develop data management systems to facilitate and implement harmonized and standardized data extraction; set up an ethical charter for future use of deidentified data; and develop a statistical analysis plan to feed the mathematical models.

This manuscript aims to describe the work accomplished by WP3 in the process of identifying and gathering datasets with anonymized individual patient-level data for secondary data analysis, map the challenges faced and propose possible solutions.

Procedures necessary to gather data

Table 1 describes the procedures necessary to gather anonymized individual patient-level data and make them publicly available.

How to identify eligible datasets?

A systematic search of peer-reviewed and grey literature (reports, thesis, and white papers) was performed to identify studies with individual patient-level data useful for PrIMAVeRa project (registered in PROSPERO, CRD42022322029) [18].

Studies were selected if reporting data for at least one of eight target AMR bacteria (carbapenem-resistant *Acinetobacter baumannii* or *Pseudomonas aeruginosa*; carbapenem- or third-generation cephalosporin-resistant *Escherichia coli* or *Klebsiella pneumoniae*; methicillin-resistant *Staphylococcus aureus*; and vancomycin-resistant *Enterococcus faecium*) and one of six infections (bloodstream, lower respiratory tract, urinary tract, intra-abdominal, skin and soft-tissue, and surgical site infections) (Table S1). These pathogens and infections were selected based on Centre for Disease Prevention and Control ranking of relevance, priority pathogen list from World Health Organization, previous AMR burden estimates, resistance prevalence and availability of vaccines or mAbs in the pipeline [19–21].

Comparator groups were patients with infection caused by susceptible isolates or without infection. Studies from high-income countries (World Bank classification) with a total sample size of at least 80 participants and at least 40 patients in the resistant group were selected (minimum sample to provide meaningful parameters for modelling).

Outcomes assessed were mortality, hospital and intensive care unit (ICU) admission, length of hospital or ICU stay, clinical or microbiological cure, infection recurrence, cognitive impairment and economic outcomes (e.g. number of outpatient and emergency room consultations, inpatient costs, ICU costs, hospital readmission). Peer-reviewed literature was performed on MEDLINE® (PubMed) and Embase (Ovid) and grey literature review used Open Grey, Database of Abstracts of Reviews of Effects, British Library Thesis Service and Clinical trial registries (Table S2) [22]. Cohort, case-control and randomized controlled studies published from 1990 to 2022 were assessed, regardless of language. Overall, 108 studies were eligible for inclusion (Table S3).

Invitation letter and eligibility questionnaire

An invitation letter for collaboration was sent by e-mail to principal investigators (PIs), senior and corresponding authors (co-authors in case of no reply) of eligible studies and co-authorship offered in publications using their datasets. Reminders sent up to five times over a 1-year period.

Table 1
Summary of procedures necessary to gather anonymized individual patient-level data and make it publicly available

Gathering data
Step 1 – Define your study protocol and minimum set of variables necessary for your study, taking in consideration different historical study designs, data types and structures.
Step 2 – Define the privacy proception level of your project and consequently the anonymization standards required; clearly describe the data anonymization techniques that should be implemented to achieve your project requirements.
Step 3 – Gather contact details from corresponding author, senior author and co-authors from the eligible studies based on your protocol definitions.
Step 4 – Send an invitation letter for collaboration to e-mail addresses identified, with follow-up reminders; offer PIs participation in secondary data analysis projects that will reuse their data. If possible, try to contact authors by phone call or in-person to strength collaboration.
Step 5 – Send a questionnaire for scientific collaboration to studies interested in sharing their data to access datasets eligibility from an ethical and legal point of view; and prepared questionnaire evaluation reports based their answers.
Step 6 – Request local EC approval, if not previously sought by the study's PI.
Step 7 – Confirm that datasets are anonymized according to your project standards, by requesting signature of an ethical charter by study's PI.
Step 8 – Sign a data access and use agreements between the study's PI and data receipt institutions legal representatives.
Step 9 – Prepare an infrastructure that PI's can use to securely transfer their anonymized datasets.
Step 10 – Standardize your metadata file using ontological terms from ICD-10, LOINC and SNOMED CT.
Step 11 – Harmonize received datasets according to your metadata file structure.
Step 12 – Securely store datasets in CDISC format in more than one server to facilitate data backup and redundancy.
Making data publicly available
Step 1 – Prepare a data management plan including information about project's long-term sustainability and description of the procedure to access data.
Step 2 – Design and implement a freely online infrastructure where information about your project, metadata file, dataset description and detailed information about datasets gathered can be publicly available for consultation.
Step 3 – Have an online application form freely available, where researchers can apply to request access to gathered data.
Step 4 – Embed clauses related to privacy and personal data and terms and conditions for participation in the application form.
Step 5 – Define a scientific committee with experts from different fields to assess data requests.
Step 6 – Manually or automatically prepare the minimal dataset based on application form information and securely transfer it to researchers.
Step 7 – Confirm that datasets are destroyed after data analysis completed.

CDISC, Clinical Data Interchange Standards Consortium; EC, ethics committee; ICD-10, International Statistical Classification of Diseases and Related Health Problems 10th Revision; LOINC, Logical Observation Identifiers Names and Codes; PI, principal investigator; SNOMED CT, Systematized Nomenclature of Medicine Clinical Terms.

Upon a positive answer, researchers were asked to complete a questionnaire to ensure studies meet the inclusion criteria from ethical and legal perspectives: type of data, data anonymization, country, informed consent and ethics committee (EC) approval, including authorization for data sharing (Table S4).

Of the 108 studies selected, no answer was obtained in 64 studies (59%). Representatives from 35 studies showed interest in collaborating, four were not interested and five have destroyed the original database. Of the 35 studies interested, 24 (69%) completed the questionnaire, of which five were excluded because of lack of individual patient-level data (Fig. 1).

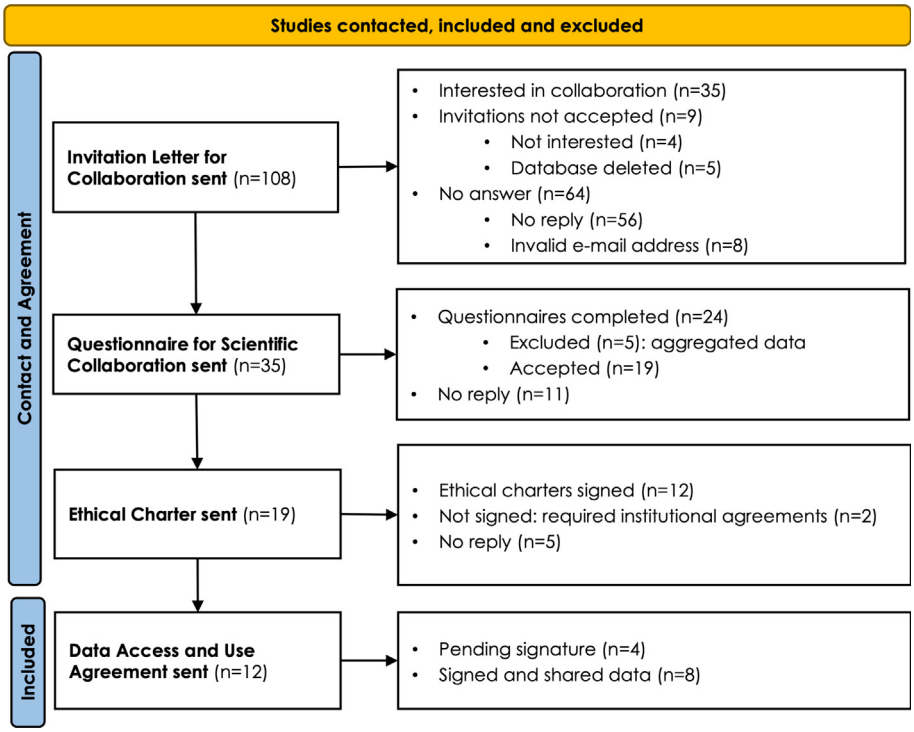

Fig. 1. Adaptation of a Preferred Reporting Items for Systematic reviews and Meta-Analysis (PRISMA) flowchart describing the procedures necessary to gather anonymized individual-level data and reasons from excluding studies in the different phases of this process.

Ethical and legal procedures

According to PrIMAVeRa project requirements, individual-level data needed to be anonymized. Although anonymized data themselves are not subject to general data protection regulation (GDPR), the process of anonymization and ensuring data cannot be re-identified requires compliance with GDPR and national regulations [23].

A law firm with expertise in data protection legislation was hired to establish the required procedures and documents in accordance with GDPR and national regulations. An ethical charter was developed to certify that: (a) all relevant authorizations from competent authorities were obtained; (b) data were obtained with the competent EC approval and informed consents signed, if needed; (c) non-personal data generated within PrIMAVeRa project could be made available through adequate public repositories; (d) all PrIMAVeRa publications would be published under open access and the PIs will be considered as co-authors; and (e) legal requirements for exporting/importing data to/from other countries were met. The ethical charter was signed by the PI as a prerequisite for signing the data access and use agreement.

The data access and use agreement, signed between PIs and PrIMAVeRa WP3 lead (Andalusian Public Health Service [SAS], Seville, Spain) authorized representatives, governs collaboration between parties in achieving PrIMAVeRa aims; guarantees that study's PI transfers anonymized individual patient data; and ensures that SAS takes all necessary technical and organizational measures to guarantee an optimal level of security, compliant with GDPR and national data protection requirements.

PrIMAVeRa WP3 study protocol 'Impact of vaccines and monoclonal antibodies on clinical outcomes in patients with infection by resistant microorganisms' was shared with the study's PIs and used in case EC approval was not historically obtained or the PI's institution required an additional approval.

Of the 19 eligible studies in the questionnaire report, 12 signed the ethical charter, two required additional specific agreements from their institutions and five did not answer. Finally, data access and use agreement was sent to 12 studies, of which eight signed it and shared their datasets by the time of this publication (Table S5); the other four were pending.

How to anonymize patient-level data?

According to the Sharing Anonymized and Functionally Effective data standard rating system, PrIMAVeRa individual patient-level data would require level 2 privacy protection, because data would be kept in a highly secure environment and could be shared externally [24]. On the basis of this, we defined the procedure for data anonymization and shared it with the PIs to ensure that datasets were correctly anonymized before sharing, which included: deletion of all direct identifiers (information that is unique to an individual, i.e. name, date of birth) and free texts (which more frequently contain direct identifiers); use of PhUSE method for dates anonymization (in which the original dates are shifted by a certain number of days, maintaining the chronological relationship between events); and 2 k-anonymization for indirect identifiers such as age, sex and country (in which each combination of indirect identifiers values appears at least twice in the dataset) [1,16,24–26].

How to harmonize and standardize datasets?

Data were semantic and syntactically harmonized according to a minimum set of variables selected by the PrIMAVeRa consortium (available in metadata file at ECRAID-Base EPI-Net; https://epi-net.eu/primavera/data-availability/) to accommodate different data types and formats, according to model's requirements. The data harmonization process was done manually to deal with different definitions and data structures; it included the translation of non-English datasets to English, mapping original study variables, rename original variables and re-categorization as needed.

To comply with FAIR principles, we use Clinical Data Interchange Standards Consortium format that aims to standardize data collection, management and reporting of trials [3,13]. Additionally, variables were mapped to ontological terms using the International Statistical Classification of Diseases and Related Health Problems 10th revision, Logical Observation Identifiers Names and Codes and Systematized Nomenclature of Medicine Clinical Terms, using the Athena tool [27].

How to provide public access to gathered data?

Public access to PrIMAVeRa WP3 data is provided by ECRAID-Base EPI-Net platform (https://epi-net.eu/primavera/about/anonymized-individual-patient-data/), managed by University of Verona (UNIVR), Verona, Italy. Metadata and dataset description files are available as well as detailed information about the original studies.

PrIMAVeRa consortium members have access to data for analysis within the scope of PrIMAVeRa project. External researchers may request access by completing an online application form including the objectives of their request, data needed, ethics and security, funding and expected scientific outputs. This form re-directs to privacy policy terms and data use terms and conditions on the ECRAID-Base EPI-Net platform to which applicants should adhere.

The requests are assessed by a scientific committee including representatives from SAS and UNIVR, and upon positive approval, requested data are securely transferred. After completion of secondary analysis, external researchers need to confirm data deletion (Fig. 2). WP3 anonymized datasets are securely stored in SAS and UNIVR servers, allowing data backup. The PrIMAVeRa data management plan includes information about project's long-term sustainability.

Challenges and possible solutions in data sharing

Table 2 describes challenges and possible solutions identified.

Researchers' interest in sharing data

We had a low response rate (41%) from study's authors even after multiple contacts and reminders. Some authors declared that data were no longer available (destroyed or lost); in other cases, the PI was retired or deceased, and other contacted co-authors declared they could not share the data. Some companies and/or institutions were also not allowed to share their data because of internal policies. For those who responded, we offered co-authorship. Previous data suggest that researchers do not receive enough credit for sharing their data and recognition, including citation of their research papers, is a critical factor to motivate data sharing [14].

Previous studies reported similar low response rate when contacting study's PIs [11,28,29]. In this sense, e-mail communication might be less efficacious than phone or in-person meetings, which are of course more complex to achieve [29]. Researchers may also be reluctant to share their data immediately after publication because this may compromise further analyses [11,13] Researchers may also be afraid of incorrect analysis and misinterpretation, highlighting the importance of their involvement in secondary data analysis [3,11,16,30] Structural difficulties related to lack of time, funding and organizational resources can also reduce researchers interest [9,13,29,31,32].

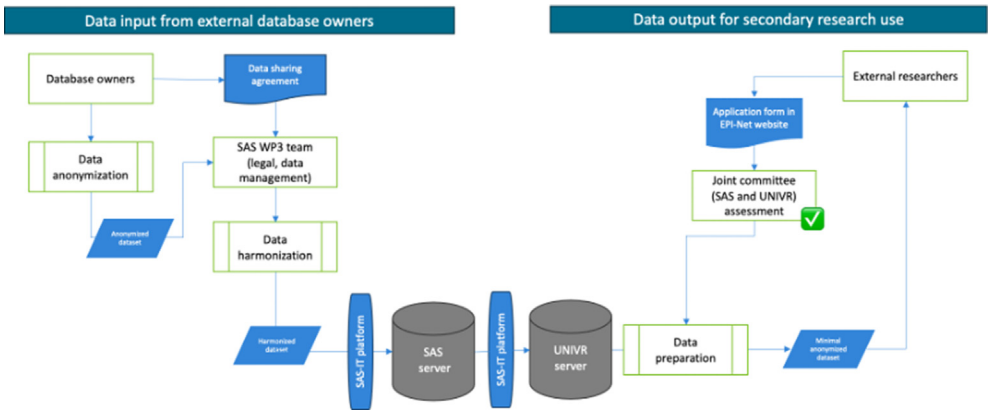

Fig. 2. PrIMAVeRa WP3 dataflow diagram. IT, information technology; PrIMAVeRa, Predicting the Impact of Monoclonal Antibodies & Vaccines on Antimicrobial Resistance; SAS, Andalusian Public Health Service; UNIVR, Verona University; WP3, work package 3.

Ethical and legal requirements

Data sharing agreements should be signed between parties and include ethical aspects for data use, patients' privacy and proper handling of data [1,3,11,16,25,30,33]. Agreements terms should be transparent, and mechanisms for reporting and handling potential breaches established [11,15,16,30]. In our experience, agreement's signature was cumbersome because of PI's difficulties in understanding legal requirements, insufficient institutional support and need of institution legal representative's signature.

There is disparity in authors' opinion about the need to obtain a specific EC approval for data sharing, even if included in the original informed consent [1,4,11,34,35]. Possible solutions can be disclosing potential data sharing mechanisms in the informed consent, sharing only anonymized data and restrain secondary data analysis within the range of original research project [25,34]. However, it is not possible to anticipate future studies that might arise; also, it is arguable that participants should have the opportunity to withhold their permission for specific studies [25].

Data management activities

Anonymization is the process of rendering personal data anonymous by dealing with direct and indirect identifiers [25,36]. It should balance the risk of reidentification and the preservation of dataset utility [10,36]. There are some platforms supporting data anonymization, but still part of the work needs to be done manually; e.g. select direct and indirect identifiers, define k-anonymization classes and anonymize dates [37,38]. Solutions to facilitate data anonymization include data collection standardization and improvement of data automation [33].

Even though data harmonization and standardization increases data interoperability and reusability in compliance with FAIR principles, their implementation can be challenging because of differences in study designs, data structures and documentation standards [3,6,13,39]. When mapping the PrIMAVeRa project variables with standardized terms, the multiplicity of terminologies was challenging, and in some cases, there were not ontological terms available that clearly described our variables, in which case the closest term was used.

Public access to data

Defining the process for public data access is complex, because different options exist ranging from free open data to access upon request. Access upon request allows more control but can be more cumbersome. To facilitate it, we developed an online application where legal clauses are embedded. When data applications are accepted, a minimal dataset is manually prepared based on project requirements and secured transferred. This manual workload could be reduced by implementing a fully automated data availability process.

In recent years, data repositories and portals emerged providing infrastructures that can store and share anonymized datasets according to funders' requirements, while providing tools for anonymization and metadata curation [2,9,11,13,40]. However, we are not aware of any unique centralized portal where researchers can deposit their data and obtain data management support; or a global tool mapping all currently available data repositories. Electronic application forms can make data access request easier, but they can also impose legal challenges because of collection of researchers' sensitive data [11,16,30]. Formal data access policies and review committees can help ensure fair and responsible data sharing [11,25].

Training and funding

Suboptimal knowledge of data sharing requirements might be overcome by training a new generation of data sharing experts [9,14,41]. Furthermore, lack of funding for data sharing makes this activity unappealing, particularly when considering the necessary resources [11,16]. Funding agencies should support including data sharing requirements in grant applications and providing specific financial support [14,41]. Despite initial costs, data value will increase with subsequent uses [2].

Strengths and limitations

Despite the challenges and barriers faced, we could gather data from eight studies, corresponding to four infections (bloodstream, urinary tract, pneumonia and intra-abdominal infections) and seven resistant pathogens (all planned except vancomycin-resistant *E. faecium*).

The main limitation of our study was the difficulty in understanding the low response rate by the PIs as we were unable to reach out many of them. Nevertheless, our results are similar to other data gathering projects, where researchers' recognition was reported as the main limitation. On the other hand, our experience dealing with the legal procedures necessary to gather anonymized individual data can be used as a toolkit in future projects. Additionally, we were able to clearly define an easy anonymization procedure that could be used when dealing with individual-level data.

Table 2
Description of challenges and possible solutions to gather anonymized individual patient data from existing studies

Challenges	Possible solutions
Researchers' interest in data sharing	
1. Corresponding author and/or PI low response rate	Subject and content in the invitation e-mail should spark author's interest to collaborate and underline potential benefits. Communication plan including multiple e-mail reminders. Contact multiple co-authors of the same study. Contact PI through network partners, phone or in-person meetings. Central database with up-to-date researchers' contacts details.
2. Lack of investigators motivation in data sharing	Recognize investigators' role, by offering co-authorship in studies that reuse their data and cite their original work. Improve culture of data sharing and reuse, by involving PI in secondary data analysis projects of their own datasets. Funding bodies should highlight the importance of data sharing by making it mandatory and providing dedicated resources.
Ethical and legal requirements	
3. EC approval for data sharing and secondary use	Informed consent should conceive information about data sharing and reuse. Restrain the secondary data analysis within the scope of the original research. Share anonymized data, to facilitate EC approval. Provide detailed protocol describing the purpose of data sharing for secondary use.
4. Compliance with GDPR	Share anonymized data that are not under the GDPR legislation. Improve literacy and favour discussions about GDPR in academic institutions.
5. Gathering data outside EU	Specific data sharing agreements should be signed between EU and non-EU institutions, due to differences in data protection legislation.
6. Clauses necessary in data sharing agreements	Sign data sharing agreement between data owners and data controllers. Data sharing agreement should describe both parties' rights and duties, and address aspects related to ethical use, participant privacy, and proper handling of data.
7. Lack of academic expertise in legal requirements for data sharing	Get support from institutions or companies with expertise in this field (e.g. data protection officers, law firms).
Data management activities	
8. Different interpretations of data anonymization	Request PIs to be responsible for anonymization of their datasets. Provide standard operating procedure for data anonymization, including easy-to-implement techniques and platforms that support data anonymization.
9. Manual workload in data harmonization	Define a minimum core set of variables essential for secondary data usage to reduce data harmonization workload. Design an electronic case report form that includes hierarchic variables to accommodate for original datasets heterogeneity (e.g. timepoints of assessment, dates, planned study follow-up). Use an interface for data gathering that provides data checks, reducing data errors.
10. Compliance with FAIR principles	Use CDISC format to standardize data collection and management, facilitating data exchange and analysis. Include ontological terms (e.g. ICD-10, LOINC and SNOMED CT) in metadata file to improve data standardization.
Public access to data	
11. Security of data storage infrastructure	Data hosting environment should comply with security standards and allow total or partial controlled data access for trustworthy researchers. Use an interface that allows batch importation, reducing workload.
12. Insufficient datasets documentation	Metadata file should be publicly available to provide an overview of dataset's content and structure. Dataset descriptor file should be publicly available and include information about study aims, study design, population under study, sample size, outcomes, etc.
13. Procedures necessary to provide public access to data	Foresee an online application form to facilitate request to data access. Facilitate signature of data sharing agreements by embed them in the application form. Establish a multidisciplinary scientific committee to assess requests taking into consideration projects scientific value and their potential to advance scientific knowledge. Reduce manual workload in data preparation by implementing a fully automated data availability process based on the application form information.
14. Uncontrolled dissemination of data shared with external researchers	Follow-up of data sharing. Request external researchers to delete and provide confirmation of dataset deletion after completion of secondary data analysis project.
15. Maintain data sustainability	Data management plan should include considerations on project's long-term sustainability, detailing procedure for public access request to gathered data. Data backup or redundancy is advised to prevent possible data breaches.
Training and funding	
16. Suboptimal knowledge of data sharing requirements	Train a new generation of data sharing experts in essential subjects such as data preparation, privacy regulations, anonymization, reporting, and governance.
17. Lack of funding for data sharing activities	Include data sharing requirements in grant applications and providing financial support for that.

CDISC, Clinical Data Interchange Standards Consortium; EC, ethics committee; EU, European Union; FAIR, findable, accessible, interoperable, reusable; GDPR, General Data Protection Regulation; ICD-10, International Statistical Classification of Diseases and Related Health Problems 10th Revision; LOINC, Logical Observation Identifiers Names and Codes; PI, principal investigator; SNOMED CT, Systematized Nomenclature of Medicine Clinical Terms.

The main strength of our project is the development of a public data repository with patient-level data on AMR infections. These data can be used not only in PrIMAVeRa studies to assess the impact of vaccines and mAbs on AMR, but also in other future studies on the same field.

Conclusion

Data sharing is important to foster research, increase transparency and reduce duplication of studies. The lack of researcher's interest in sharing their data is one of the main bottlenecks when

making data publicly available. Insufficient funding for data sharing and scarce legal and data managing support may be some of the barriers that make researchers unwilling to share their data. Previous studies have reported researchers' recognition as the main factor that restrains researchers from data sharing.

Author contributions

M.G. contributed to the conception and design of the work, systematic search execution, data acquisition, literature review, drafted the work and approved the submitted version. A.d.l.S.B. contributed to the conception and design of the work, data acquisition, literature review, drafted the work and approved the submitted version. E.R.-M., L.B.P., and V.P. contributed to the conception and design of the work, data acquisition and approved the submitted version. E.S.-R. and J.R.-B. contributed to the conception and design of the work, substantial revised the work and approved the submitted version. N.H.-K. and M.E.A.d.K. contributed to the conception and design of the work, systematic search protocol and execution and approved the submitted version. A.P., Q.J.L., E.A., V.V., A.D., J.V.R., A.P., F.A., R.J.D., and E.T contributed to the conception and design of the work and approved the submitted version. All authors agree to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved and the resolution documented in the literature. M.G. and A.d.l.S.B. contributed equally to this work. E.S.-R. and J.R.-B. contributed equally to this work.

Transparency declaration

Potential conflict of interest

V.V. is a GSK staff and owns GSK shares. The other authors report no conflicts of interest relevant to this article.

Financial report

This work has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 101034420 (PrIMAVeRa). This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation program and European Federation of Pharmaceutical Industries and Associations (EFPIA). This communication reflects the author's view and neither Innovative Medicines Initiative (IMI) nor the European Union, EFPIA, or any associated partners are responsible for any use that may be made of the information contained herein. This work has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 965313 (ECRAID-Base).

Data availability

Anonymized individual patient-level data are available upon request in the ECRAID-Base EPI-Net platform (https://epi-net.eu/primavera/about/anonymized-individual-patient-data/).

Acknowledgements

We acknowledge the work performed by Alberto Moreno Conde, Marta Cuaresma Giráldez and Elena Perez Ornedo, from Andalusian Public Health Service Innovation team, in developing the infrastructure to securely store Predicting the Impact of Monoclonal Antibodies & Vaccines on Antimicrobial Resistance data and helping in mapping variables to standardized ontological terms. We acknowledge Corinne Willame from Janssen Pharmaceutical Companies of Johnson & Johnson for reviewing the final version of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cmi.2025.01.024.

References

[1] Committee on Strategies for Responsible Sharing of Clinical Trial Data. Board on Health Sciences Policy, Institute of Medicine. Sharing clinical trial data: maximizing benefits, minimizing risk. Appendix B, concepts and methods for de-identifying clinical trial data. Sharing clinical trial data: maximizing benefits, minimizing risk. National Academies Press (US); 2015.
[2] Loder E, Groves T. The BMJ requires data sharing on request for all trials. BMJ 2015;350:h2373. https://doi.org/10.1136/bmj.h2373.
[3] Fletcher C, Hollis S, Burger HU, Gerlinger C. Statistical guidance for responsible data sharing: an overview. BMC Med Res Methodol 2016;16(Suppl 1):74. https://doi.org/10.1186/s12874-016-0168-5.
[4] Taichman DB, Sahni P, Pinborg A, Peiperl L, Laine C, James A, et al. Data sharing statements for clinical trials: a requirement of the International Committee of Medical Journal. Lancet 2017;389:e12–4. https://doi.org/10.1016/S0140-6736(17)31282-5.
[5] European Medicines Agency. Clinical data publication. EMA Clinical data publication; 2024 [Internet]. 2024 [2024 February 22]. Available from: https://www.ema.europa.eu/en/human-regulatory-overview/marketing-authorisation/clinical-data-publication.
[6] Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, et al. The FAIR guiding principles for scientific data management and stewardship. Sci Data 2016;3:160018. https://doi.org/10.1038/sdata.2016.18.
[7] Digital Science, Goodey G, Hahnel M, Zhou Y, Jiang L, Chandramouliswaran I, et al. The state of open data 2022. 2022. https://doi.org/10.6084/m9.figshare.21276984.v5.
[8] Krumholz HM, Waldstreicher J. The Yale Open Data Access (YODA) project—a mechanism for data sharing. N Engl J Med 2016;375:403–5. https://doi.org/10.1056/NEJMp1607342.
[9] Mansmann U, Locher C, Prasser F, Weissgerber T, Sax U, Posch M, et al. Implementing clinical trial data sharing requires training a new generation of biomedical researchers. Nat Med 2023;29:298–301. https://doi.org/10.1038/s41591-022-02080-y.
[10] Coady SA, Wagner E. Sharing individual level data from observational studies and clinical trials: a perspective from NHLBI. Trials 2013;14:201. https://doi.org/10.1186/1745-6215-14-201.
[11] Hopkins C, Sydes M, Murray G, Woolfall K, Clarke M, Williamson P, et al. UK publicly funded Clinical Trials Units supported a controlled access approach to share individual participant data but highlighted concerns. J Clin Epidemiol 2016;70:17–25. https://doi.org/10.1016/j.jclinepi.2015.07.002.
[12] Merson L, Gaye O, Guerin PJ. Avoiding data dumpsters—toward equitable and useful data sharing. N Engl J Med 2016;374:2414–5. https://doi.org/10.1056/NEJMp1605148.
[13] Sudlow R, Branson J, Friede T, Morgan D, Whately-Smith C. EFSPI/PSI working group on data sharing: accessing and working with pharmaceutical clinical trial patient level datasets—a primer for academic researchers. BMC Med Res Methodol 2016;16(Suppl 1):73. https://doi.org/10.1186/s12874-016-0171-x.
[14] Digital Science, Hahnel M, Smith G, Schoenenberger H, Scaplehorn N, Day L. The state of open data 2023. Digital Science. 2023. https://doi.org/10.6084/m9.figshare.24428194.v1.
[15] Navar AM, Pencina MJ, Rymer JA, Louzao DM, Peterson ED. Use of open access platforms for clinical trial data. JAMA 2016;315:1283–4. https://doi.org/10.1001/jama.2016.2374.
[16] Sydes MR, Johnson AL, Meredith SK, Rauchenberger M, South A, Parmar MK. Sharing data from clinical trials: the rationale for a controlled access approach. Trials 2015;16:104. https://doi.org/10.1186/s13063-015-0604-6.
[17] PrIMAVeRa. PrIMAVeRa project [Internet]. 2022 [2023 November 5]. Available from: https://www.primavera-amr.eu.
[18] Guedes M, Rodríguez-Baño J. Clinical outcomes in patients with infection by resistant microorganisms [Internet]. 2022 [2024 November 25]. Available from: https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=322029.
[19] Cassini A, Högberg LD, Plachouras D, Quattrocchi A, Hoxha A, Simonsen GS, et al. Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European economic area in 2015: a population-level modelling analysis. Lancet Infect Dis 2019;19:56–66. https://doi.org/10.1016/S1473-3099(18)30605-4.
[20] European Centre for Disease Prevention and Control. Antimicrobial resistance in the EU/EEA (EARS-Net)—annual Epidemiological Report 2019 [Internet]. Stockholm: ECDC; 2020. 2020 [2024 November 25]. Available from: https://www.ecdc.europa.eu/sites/default/files/documents/AER-EARS-Net%202019-final.pdf.

[21] Frost I, Sati H, Garcia-Vello P, Hasso-Agopsowicz M, Lienhardt C, Gigante V, et al. The role of bacterial vaccines in the fight against antimicrobial resistance: an analysis of the preclinical and clinical development pipeline. Lancet Microbe 2023;4:e113–25. https://doi.org/10.1016/S2666-5247(22)00303-2.
[22] Hassoun-Kheir N, Guedes M, Ngo Nsoga MT, Argante L, Arieti F, Gladstone BP, et al. A systematic review on the excess health risk of antibiotic-resistant bloodstream infections for six key pathogens in Europe. Clin Microbiol Infect 2024;30(Suppl 1):S14–25. https://doi.org/10.1016/j.cmi.2023.09.001.
[23] European Parliament and of the Council. General data protection regulation (GDPR). https://gdpr-info.eu/; 2016.
[24] Bamford S, Lyons S, Arbuckle L, Chetelat P. Sharing Anonymized and Functionally Effective (SAFE) data standard for safety sharing rich clinical trial data. Appl Clin Trial. 2022;31:30–43.
[25] Malin B, Karp D, Scheuermann RH. Technical and policy approaches to balancing patient privacy and data sharing in clinical and translational research. J Investig Med 2010;58:11–8. https://doi.org/10.2310/JIM.0b013e3181c9b2ea.
[26] phuse. PhUSE data de-identification standard for SDTM 3.2. Appendix 1. Date Offsetting; 2015. https://phuse.s3.eu-central-1.amazonaws.com/Deliverables/Data+Transparency/Data+De-Identification+Standard+for+SDTM+3.2+%E2%80%93+Appendix+1%3A+Date+Offsetting+Version+1.0.pdf.
[27] Athena. 2015 [2024 March 6]. Available from: https://athena.ohdsi.org/search-terms/start.
[28] Esmail LC, Kapp P, Assi R, Wood J, Regan G, Ravaud P, et al. Sharing of individual patient-level data by trialists of randomized clinical trials of pharmacological treatments for COVID-19. JAMA 2023;329:1695–7. https://doi.org/10.1001/jama.2023.4590.
[29] Ventresca M, Schünemann HJ, Macbeth F, Clarke M, Thabane L, Griffiths G, et al. Obtaining and managing data sets for individual participant data meta-analysis: scoping review and practical guide. BMC Med Res Methodol 2020;20:113. https://doi.org/10.1186/s12874-020-00964-6.
[30] Alfonso F, Adamyan K, Artigou JY, Aschermann M, Boehm M, Buendia A, et al. Data sharing: a new editorial initiative of the international Committee of Medical Journal. In: Implications for the editors' network. Kardiol Pol, vol. 75; 2017. p. 512–7. https://doi.org/10.5603/KP.2017.0086.
[31] de Carvalho ECA, Batilana AP, Simkins J, Martins H, Shah J, Rajgor D, et al. Application description and policy model in collaborative environment for sharing of information on epidemiological and clinical research data sets. PLoS One 2010;5:e9314. https://doi.org/10.1371/journal.pone.0009314.
[32] Modjarrad K, Moorthy VS, Millett P, Gsell PS, Roth C, Kieny MP. Developing global norms for sharing data and results during public health emergencies. PLOS Med 2016;13:e1001935. https://doi.org/10.1371/journal.pmed.1001935.
[33] Tucker K, Branson J, Dilleen M, Hollis S, Loughlin P, Nixon MJ, et al. Protecting patient privacy when sharing patient-level data from clinical trials. BMC Med Res Methodol 2016;16(Suppl 1):77. https://doi.org/10.1186/s12874-016-0169-4.
[34] Zink A, Silman AJ. Ethical and legal constraints on data sharing between countries in multinational epidemiological studies in Europe report from a joint workshop of the European League against Rheumatism standing committee on epidemiology with the 'AutoCure' project. Ann Rheum Dis 2008;67:1041–3. https://doi.org/10.1136/ard.2007.080598.
[35] Menikoff J. Letter from Jerry Menikoff, MD, JD, Director, Office for Human Research Protections, to ICMJE Secretariat. 2017 [Internet]. 2017 [2023 December 12]. Available from: https://icmje.org/news-and-editorials/menikoff_icmje_questions_20170307.pdf.
[36] Agencia Española Protección Datos. European Data Protection Supervisor. 10 misunderstandings related to anonymization. 2021 [Internet]. 2021. Available from: https://edps.europa.eu/system/files/2021-04/21-04-27_aepd-edps_anonymisation_en_5.pdf.
[37] ARX. ARX—data anonymization tool. https://arx.deidentifier.org/.
[38] OpenAIRE. Amnesia anonymization tool. https://amnesia.openaire.eu/.
[39] Fortier I, Raina P, Van den Heuvel ER, Griffith LE, Craig C, Saliba M, et al. Maelstrom Research guidelines for rigorous retrospective data harmonization. Int J Epidemiol 2017;46:103–5. https://doi.org/10.1093/ije/dyw075.
[40] Maxwell L, Shreedhar P, Dauga D, McQuilton P, Terry RF, Denisiuk A, et al. FAIR, ethical, and coordinated data sharing for COVID-19 response: a scoping review and cross-sectional survey of COVID-19 data sharing platforms and registries. Lancet Digit Health 2023;5:e712–36. https://doi.org/10.1016/S2589-7500(23)00129-2.
[41] Tacconelli E, Gorska A, Carrara E, Davis RJ, Bonten M, Friedrich AW, et al. Challenges of data sharing in European COVID-19 projects: a learning opportunity for advancing pandemic preparedness and response. Lancet Reg Health Eur 2022;21:100467. https://doi.org/10.1016/j.lanepe.2022.100467.