IMS Health & Quintiles are now IQVIA™

White paper

MANAGING PRIVACY

The imperative real-world evidence capability

KHALED EL EMAM, PhD, President, Privacy Analytics, IQVIA
BENJAMIN HUGHES, PhD, Global Head of Technology, Real-World Insights, IQVIA

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Increasing leverage of electronic health records (EHRs), imaging systems and social media is delivering groundbreaking patient-centric insight. However, as companies seek deeper understanding of health outcomes via patient-level data, privacy-enhancing capabilities are becoming a critical source of leadership – and key to unlocking the full value of real-world evidence (RWE).

BETTER RESEARCH DEMANDS BETTER PRIVACY

Secondary use of individual-level patient data for health research has unparalleled potential to improve healthcare quality and drive medicines innovation, benefiting individual patients and society as a whole. To optimize its value for scientific research and meet tightening privacy regulations, there is a pressing need for a systematic approach to privacy management and de-identification of data. Life sciences companies can take a lead in implementing best practice using risk-based privacy-enhancing techniques.

GREATER RWE SOPHISTICATION, GREATER PRIVACY CONSIDERATION

Pharma is accessing an unprecedented depth and breadth of clinical data. Leading companies have built RWE platforms across multiple countries, encompassing vast real-world data (RWD) collections and analytics technology. The variety of RWD has exploded, spanning hundreds of databases or registries, including data directly sourced from providers, and extensive data linkage. This significantly expands research potential, but also creates elevated re-identification risk.

FROM BLUNT TOOLS TO RISK-BASED PRIVACY TECHNIQUES

Current de-identification methods, and commonly used data masking in particular, do not appropriately address ever increasing RWE data sophistication nor changing privacy regulation. RWE leadership, therefore, requires privacy leadership and the application of new standards in risk-based privacy. Use of risk-based de-identification software, which enables holistic overall privacy governance frameworks, both accelerates RWE strategies and ensures a continuous compliant flow of RWD.

EXTENDED OPPORTUNITIES FROM PRIVACY LEADERSHIP IN RWE

Risk-based privacy techniques can be applied systematically in a relatively short period of time to achieve a step-change in RWE capabilities. Their adoption also opens up new opportunities for companies to partner in knowledge and data sharing with health systems, which are also embracing risk-based approaches, to collectively work towards the ultimate goal of improving health outcomes.

Current de-identification methods, and commonly used data masking in particular, do not appropriately address ever increasing RWE data sophistication nor changing privacy regulation.

BETTER RESEARCH DEMANDS BETTER PRIVACY

Levering patient-level Protected Health Information (PHI) is essential for scientific research. Originating from various sources, including EHRs, PHI plays a vital role in the development of innovative medicines and improvement of health system performance. A number of high profile initiatives, including PCORI[i] in the USA and IMI[ii] in Europe, seek to maximize this potential – a point underscored by the American Society of Clinical Oncology (ASCO): "When trial data inform our decisions, we tap into only 3% of the cancer patient population.… To improve care for every patient, we need insights from the other 97% of people receiving cancer care."[1]

The ethical and moral case for facilitating the use of PHI for research and the overall benefit of society is compelling, even without consent which is often not practical to obtain.[2] Various regulations, from HIPAA[iii] in the USA to the recent changes in European data protection law, recognize this, and data may be utilized without explicit consent when in the public interest.[3] Additionally, public attitudes, especially among individuals with higher awareness and understanding of EHRs, are supportive of use without explicit consent via de-identified records.[4]

However, two barriers exist to effectively de-identify data. First, there is ambiguity in data protection regulations in selected countries, and even when clear, these are not always implemented in a systematic and apolitical manner.[5] Second, the de-identification approaches applied need to avoid destroying the value of the data for research.[3] Both of these challenges may best be addressed through capabilities in information technology, which now exist at global scale.

When trial data inform our decisions, we tap into only 3% of the cancer patient population.... To improve care for every patient, we need insights from the other 97% of people receiving cancer care.[1]

GREATER RWE SOPHISTICATION, GREATER PRIVACY CONSIDERATION

The ever increasing electronic capture of healthcare data is a core driver in the expanding business impact of RWE, enabling companies to improve decision making across drug development. Through comprehensive data platforms that link information from multiple sources, such as EHRs, claims databases, prescription data, lab results and social media networks, they are gaining an unparalleled 360-degree view of treatment patterns and patient outcomes. However, even leading companies with established comprehensive RWE networks and data platforms are only beginning to realize the full potential, worth billions of dollars to Pharma.[6]

RWE efforts initially focused on claims and pharmacy datasets, with limited data variety and scenarios to uniquely identify a particular individual. In these cases, basic masking techniques were sufficient to de-identify, using suppression to remove key identifiers.

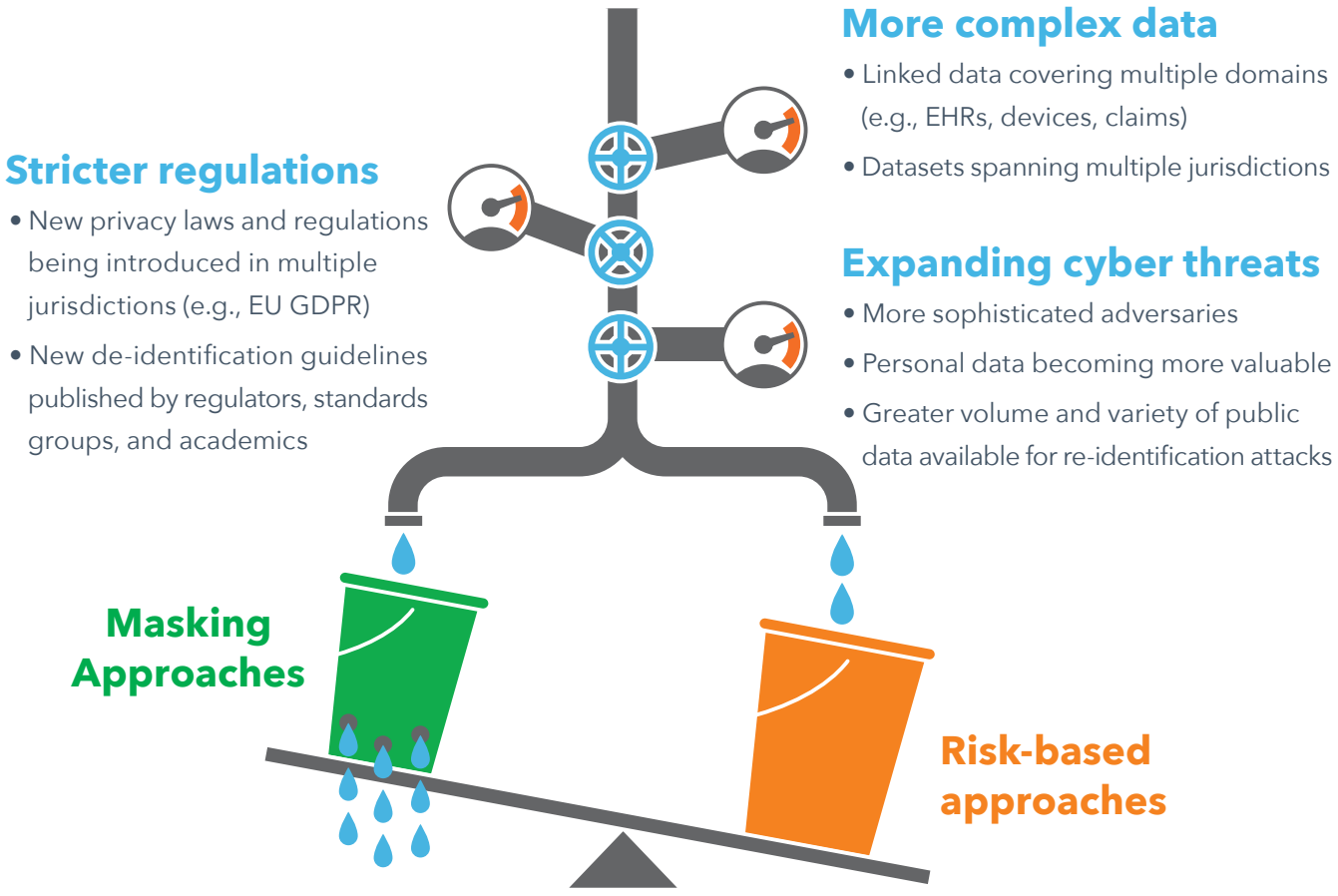
Today, as companies build access to a far greater volume and variety of RWD, there is potential for a high degree of sensitive information to be processed, which, if not dealt with correctly, could impact patient privacy. Executives driving RWE strategies must pay close attention to privacy compliance or run the risk of exposing their organization to the increasing financial penalties associated with breaches, as well as consequent reputational damage.

i. Patient-Centered Outcomes Research Institute. http//www.pcori.org
ii. Innovative Medicines Initiative. https://www.imi.europa.eu/
iii. Health Insurance Portability and Accountability Act. https://www.hhs.gov/hipaa/for-professionals/privacy/

The answer to mastering privacy does not lie in more use of masking, which either negates the analytic potential of data or fails to ensure sufficient anonymity. Masking all indirect or quasi-identifiers, for example, such as age, gender, geographic location or specific moments in time, destroys utility for healthcare research. Conversely, not masking other potentially sensitive information, such as dates or rare diagnoses, means it may be possible to re-identify a person using a unique set of values.

Indeed, the combination of more complex data, stricter privacy regulations and de-identification guidelines, and increasing availability of public data and sophistication in cyber attacks that can lever that data, have all exposed commonly used masking or typical home-grown internal privacy techniques as antiquated methods and tools (see Figure 1). It is only through more sophisticated, risk-based privacy-enhancing techniques which produce high quality de-identified data, that healthcare organizations can both meet new privacy standards such as the EU General Data Protection Regulation (GDPR)[7], and unlock the full potential of groundbreaking RWE.

Figure 1: From Masking To Risk-Based Privacy Techniques



FROM BLUNT TOOLS TO RISK-BASED PRIVACY TECHNIQUES

Risk-based methodologies (sometimes known as Expert Determination in the USA) enable de-identified and useful data for healthcare research via statistical techniques that minimize the probability of re-identification. Such techniques, including k-anonymity,[8] have emerged through a large body of academic research and offer particular advantages in privacy protection as follows:

- **Increased quality, utility and agility.** Risk-based methods enable more agile approaches, allowing high-risk chronological or geo-spatial information to be preserved. Dates, for example, can be generalized or aggregated rather than suppressed entirely, or zip codes can be shifted. This not only improves the utility of the data for research but also supports dynamic trade-offs; if geo-analysis is surplus to the research question, such sensitive information can be forsaken and other granular data retained.
- **Global relevance.** Internationally respected industry associations, including the Health Information Trust Alliance (HITRUST), Institute of Medicine (IOM), Canadian Council of Academies, European-based PhUSE, and HIPAA regulators, all recommend risk-based methods, ensuring appropriate guidance in a global environment.[iv]
- **Application beyond RWE.** Risk-based methods bridge the data divide (e.g., being recommended by the European Medicines Agency (EMA) for clinical trial transparency), enabling companies to take a uniform approach across different data types.

Privacy requirements are many and varied, and an increasing challenge for companies accessing ever growing numbers of diverse datasets. Risk-based approaches enable consistency in governing privacy by allowing the same methodology to be applied systemically across multiple complex scenarios. Companies that establish a holistic information governance framework can advance their RWE strategy with continuous automated flows of RWD, confident that all appropriate privacy measures are in place. They can further be positioned to collaborate with other stakeholders embracing risk-based privacy management, to share data and knowledge as part of a learning health system. This both strengthens their own evidence base and enhances their ability to advance health outcomes at a broader level.

DEVELOPING A PRIVACY GOVERNANCE FRAMEWORK FOR ACCELERATED RWE

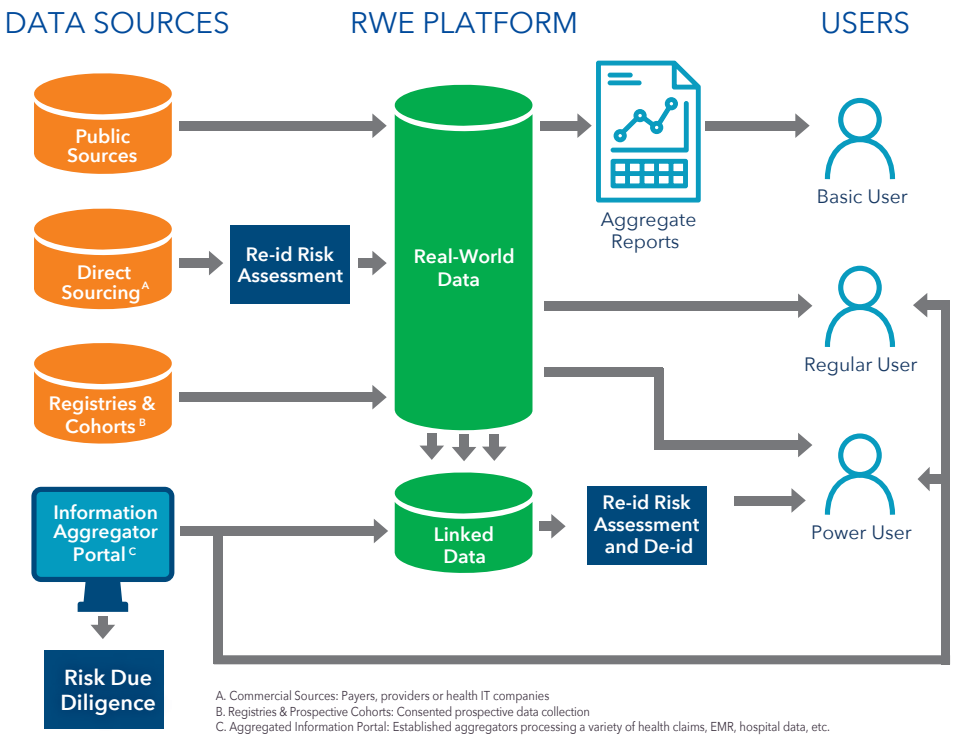
Holistic RWE information governance frameworks are established by segmenting different incoming and outgoing data flows for consumption by healthcare researchers, and applying the appropriate risk-based methods for each type.

As shown in Figure 2, companies typically access RWD from four major sources:

1. **Public sources,** such as epidemiologically-based statistics or census data
2. **Direct sourcing,** through partnering with health plans or health systems
3. **Registries & cohorts,** often generated prospectively and with patient consent
4. **Information aggregator portals** (e.g., IQVIA E360™) providing access to de-identified datasets

iv. See Appendix for full list of organizations endorsing risk-based techniques and methods

Figure 2: Illustrative RWE Privacy Govenance Framework



ASSESSMENT AND IMPLEMENTATION

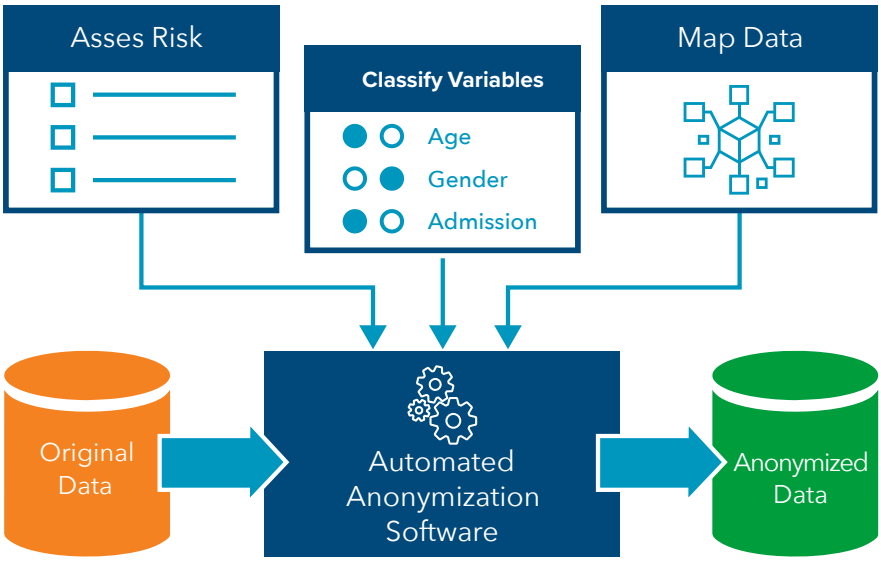
For incoming flows, the priority is to assess the risk of directly sourced data, for example from specialized commercial vendors or via collaborations with different health systems. It is not necessary at this stage to assess other datasets; public data is already in the public domain, and cohorts and registries generated through consent do not yet need to be de-identified. In the case of data from established information aggregators with data distribution portals and privacy policies, the immediate requirement is to perform due diligence on their privacy measures (see Figure 2).

The growing possibility of linkage across datasets means that further assessments and de-identification may be required once the data is linked, given the greater risk of re-identification. Increasingly, companies are keen to use consented data, such as clinical trial or registry cohorts, in conjunction with other RWD, which also requires the resulting dataset to be fully de-identified.

Implementation starts by assessing the privacy risk of each flow. A critical step is defining the usage context: who has data access; under what security and privacy controls; specific sensitive elements of the dataset; and intended use. The context is combined with statistical analysis of the data contents to determine the required de-identification routines, and ensure anonymity while maintaining the highest data quality and usability (see Figure 3, next page).

Increasingly, companies are keen to use consented data, such as clinical trial or registry cohorts, in conjunction with other RWD, which also requires the resulting dataset to be fully de-identified.

Figure 3: An Automated Date De-Identification System



FAST EVALUATION, MINIMIZED RISK AND BEST PRACTICE APPLICATION

Each major type of data flow or internal consumer – from the data source owner, through the internal RWE platform, to different types of internal healthcare researchers – can be quickly assessed using risk-based techniques. It can then be deployed on regular and recurring flows (e.g., weekly, monthly or quarterly) across millions of de-identified records to hundreds of potential internal consumers.

Assessing data flows is detailed work. However, with adequate training and knowledge companies can rapidly evaluate each flow. Within months they can be confident they are both minimizing their risk and demonstrating the systematic application of best practices in compliance with privacy legislation, while dramatically accelerating their RWE strategy.

EXTENDED OPORTUNITIES THROUGH LEARNING HEALTH SYSTEMS

Various healthcare stakeholders and governments advocate learning health systems[9] with their ability to progress healthcare through knowledge sharing. PHI from millions of people can rapidly advance the evidence base for clinical care. It can fill major gaps in understanding of healthcare costs, the benefits and risks of drugs and procedures, and geographic variations, or enable personalized medicine.

One such learning health system is CancerLinQ,[10] an initiative of ASCO. This is first and foremost a quality measurement and reporting system, allowing oncologists to harness the depth and power of their patients' clinical records to improve care delivery.[v] Having triggered important questions about the use of big data in healthcare, CancerLinQ is now required to provide oncologists with up-to-date access to de-identified EHR data from millions of cancer patients to enable more personalized treatment plans. Risk-based methods were deployed to de-identify these records before adding them to the CancerLinQ portal, ensuring a continuous flow of RWD between oncologists, with no compromise on privacy.

As more health systems embrace risk-based privacy methods for de-identification, broader opportunities open up for companies leveraging these approaches to partner, share data and directly participate in learning health systems to drive improved patient outcomes.

TAKING THE LEAD IN RWE PRIVACY MANAGEMENT

Increasingly, the onus of RWE leadership in Pharma lies with a select number of senior executives responsible for systematically building dedicated RWE capabilities Working together with the head of privacy or compliance and commercial franchise leads, they can act to accelerate their RWE strategy using risk-based privacy and industrialized software platforms by:

- **Baselining key RWE datasets** across the organization and users of this data
- Creating an **RWE governance framework** for different departments and data types
- Evaluating **risk-based privacy software** and engaging RWE technical leads on application
- **Collaborating with third parties,** including health systems partners, on implementation
- Developing a final **execution plan to deploy technologies**

Such a journey may seem daunting, but in assisting companies to build their expertise through the IQVIA Privacy Analytics De-Id University[vi] or senior leadership workshops, we have seen them quickly develop a compelling program. Over a period of only 12 months, they can create a step-change in their RWE capabilities and work more effectively with health systems to understand and improve patient outcomes.

v. https://www.privacy-analytics.com/files/Asco-Case-Study.pdf
vi. IQVIA Privacy Analytics De-id University at http://www.privacy-analytics.com/de-id-university/

As more health systems embrace risk-based privacy methods for de-identification, broader opportunities open up for companies leveraging these approaches to partner in data sharing via direct participation in learning health systems to drive improved patient outcomes.

APPENDIX

ORGANIZATIONS AND THEIR STANDARDS ENDORSING RISK-BASED METHODS

Canadian Institute for Health Information in collaboration with Canada Health Infoway:
'Best practice' Guidelines for Managing the Disclosure of De-identified Health Information

Council of Canadian Academies:
Accessing Health and Health-related Data in Canada

European Medicines Agency:
External guidance on the implementation of the European Medicines Agency policy on the publication of clinical data for medicinal products for human use

Health Information Trust Alliance:
HITRUST De-Identification Framework

Institute of Medicine:
Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk

PhUSE (Pharmaceutical Users Software Exchange):
De-Identification Standards for CDISC SDTM 3.2

Ontario Privacy Commissioner:
De-identification Guidelines for Structured Data

UK Anonymization Network:
The anonymization decision making framework

UK Information Commissioner's Office:
Anonymisation: managing data protection risk code of practice

US Department of Health and Human Services:
Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule

GLOSSARY OF PRIVACY TERMS

De-identification:
The application of expert determination and/or k-anonymity, rather than masking, such that the dataset is considered truly de-identified

Direct identifiers:
The fields within a dataset that can easily be used alone to uniquely identify individuals. Examples include name, health card number, credit card number or email address

Expert determination:
The use of statistical probabilities to oversee the de-identification process to ensure very low risk of re-identification

K-anonymity:
Data such that the information for each person cannot be distinguished from a minimum set of other individuals in the dataset

Masking:
A widely used approach to suppress identifiers, which can destroy data utility or not fully ensure anonymity in complex healthcare data

Quasi-identifiers:
The fields within a dataset that can be used in combination with one another to identify individuals. Examples include birth date, sex, marital status, ethnic origin or post/zip code

REFERENCES

1. https://cancerlinq.org
2. Porsdam Mann S, Savulescu J, Sahakian BJ. 2016 Facilitating the ethical use of health data for the benefit of society:electronic health records, consent and the duty of easy rescue. Phil. Trans. R. Soc. A 374: 20160130. http://dx.doi.org/10.1098/rsta.2016.0130
3. Quinn P. The Anonymisation of Research Data – A Pyric Victory for Privacy that Should Not Be Pushed Too Hard by the EU Data Protection Framework? European Journal of Health Law, 2016, July; 10.1163/15718093-12341416
4. Riordan F, Papoutsi C, Reed JE, Marston C, Bell D, Majeed A. Patient and public attitudes towards informed consent models and levels of awareness of Electronic Health Records in the UK. Int J Med Inform, 2015; 84(4): 237–247. doi: 10.1016/j.ijmedinf.2015.01.008. Epub 2015 Jan 20
5. Hopf YM, Francis J, Helms PJ, Haughney J, Bond C. Core requirements for successful data linkage: an example of a triangulation method. BMJ Open 2016; 6: e011879 doi:10.1136/bmjopen-2016-011879
6. Hughes B, Kessler M, McDonell A. Breaking New Ground with RWE: How Some Pharmacos are Poised to Realize a $1 Billion Opportunity. IMS Health White Paper, August, 2014
7. Stalla-Bourdillon S, Knight A. Anonymous data v. personal data—a false debate: an EU perspective on anonymization, pseudonymization and personal data. Wisconsin International Law Journal, 1-38. 2016
8. Privacy Analytics. De-identification 201: Fundamentals of data de-identification. Available at: http://www.privacy-analytics.com/de-id-university/white-papers/de-identification-201/. August, 2015
9. Etheredge LM. A Rapid-Learning Health System. Health Affairs, March, 2007; 26(2): w107-w118
10. Sledge GW, Hudis CA, Swain SM, Yu PM, Mann JT, Hauser RS, Lichter AS. ASCO's Approach to a Learning Health Care System in Oncology. J Oncol Pract., 2013, May; 9(3): 145-148. doi: 10.1200/JOP.2013.000957

AUTHORS

Dr. Khaled El Emam
President
Privacy Analytics
IQVIA

Dr. Benjamin Hughes
Global Head of Technology
Real-World Insights
IQVIA

RELATED READING AND RECENT WORKS

- El Emam K, Arbuckle L. Anonymizing Health Data: Case studies and Methods to Get You Started. 1st Ed. Sebastopol, CA: O'Reilly Media Inc; 2013
- Porsdam Mann S, Savulescu J, Sahakian BJ. Facilitating the ethical use of health data for the benefit of society:electronic health records, consent and the duty of easy rescue. 2016 Phil. Trans. R. Soc. A 374: 20160130. http://dx.doi.org/10.1098/rsta.2016.0130
- Forrester's 2015 Data Privacy Heat Map. Forrester. October 13, 2015
- Basu A, Axelsen K, Grabowski DC, Meltzer DO, Polsky D, R DB, Wiederkehr D, Philipson TJ Real-World Data. Policy Issues Regarding their Access and Use. Medical Care, 2016 Dec; 54(12): 1038-1044
- El Emam K. Guide to the De-Identification of Personal Health Information. 1st Edition. Boca Raton, FL: CRC Press/Auerbach Publications; 2013
- Meunier M-A, Tirosh A. Market Guide for Data Masking. Gartner. 6 February 2017. G00312999

ABOUT IQVIA

IQVIA Real-World Insights (RWI) team helps clients build powerful platforms around real-world data (RWD) to provide leading-edge clinical and commercial insights to increase organizational performance - by tapping into uniquely rich patient centric insights that support R&D, HEOR, drug safety, pharmacoepidemiology and commercial teams. In doing so, IQVIA aims to adhere to the highest legal, ethical and privacy standards.[vii]

IQVIA does not provide legal advisory services, assistance or opinions of any kind, and this white paper does not represent legal advice or recommendations on privacy and legal compliance. The statements in this white paper are reflective of, and cite, various organizations and individuals' opinions on healthcare data privacy, rather than IQVIA company policy on data privacy or the specific opinions of IQVIA executive officers.

ABOUT PRIVACY ANALYTICS

Privacy Analytics enables healthcare organizations to quickly and easily apply a risk-based responsible de-identification methodology that ensures individual privacy and legal compliance. Privacy Analytics is the only company to offer expert training, software, peer-reviewed methodology and valued-added services that protect the privacy of individuals while enabling organizations to share data for secondary purposes.

Privacy Analytics' software is compliant with regulations and globally accepted standards and guidelines, including those from the Institute of Medicine (IOM), Health Information Trust Alliance (HITRUST), PhUSE, the Council of Canadian Academies, as well as HIPAA and the EU General Data Protection Regulation. Additional information is available at www.privacy-analytics.com

vii. https://www. www.iqvia.com/about/privacy-commitment

Contact us
RWIinfo@iqvia.com
iqvia.com/contactus

IMS Health & Quintiles are now