



Public Biological Databases and the *Sui Generis* Database Right

Alexander Bernier · Christian Busse · Tania Bubela

Accepted: 31 July 2023 / Published online: 29 August 2023
© The Author(s) 2023

Abstract The *sui generis* database right is an intellectual property right created in the European Union to stimulate investment in the curation of databases. Since its inception, communities engaged in research and development efforts have questioned its potential to incentivise database production, and posit that it stifles productive downstream uses of existing datasets. European courts have restricted the right's ambit through a restrictive interpretation of the circumstances in which it applies, which we argue, enables downstream use of biological databases. Nonetheless, residual ambiguities about potential infringement of the right exist. The prospect of unintentional infringement can frustrate downstream innovation. These ambiguities are compounded because the criteria that determine whether or not the right applies are reliant on information that is not available to the prospective downstream users of public datasets. Repealing the *sui generis* database right is recommended. Legislatures are advised to refrain from the implementation of broad novel intellectual property rights in the future, without first adopting safeguards that

The authors wish to thank Daniel Emerling, Susanna Marquez, Jamie Scott, and Jacob S. Sherkow for their significant substantive contributions to the development of the manuscript. Alexander Bernier further wishes to thank Marc Bourreau, Zichuan Li and Emmanuel Raynaud for their thoughtful comments and input on the substance, structure, and overall articulation of an earlier draft of this manuscript.

A. Bernier (✉)

Ph.D. student; Faculty of Law, University of Toronto, Toronto, Canada
e-mail: alexander.bernier@mail.mcgill.ca

C. Busse

Ph.D.; Team Leader, German Cancer Research Center, Heidelberg, Germany
e-mail: christian.busse@dkfz-heidelberg.de

T. Bubela

Ph.D.; Professor and Dean, Faculty of Health Sciences, Simon Fraser University, Burnaby, Canada
e-mail: fhsdean@sfu.ca

mitigate the potential for such rights to frustrate the reuse of available intangibles to the detriment of pro-social innovation.

Keywords Innovation policy · Open science · Public biological databases · *Sui generis* database right

1 Introduction

Public databases have become a backbone of modern data-driven biological research. Especially in the European Union (EU), it is expected that large-scale research data infrastructure projects such as the European Open Science Cloud (EOSC) will expand existing databases over the coming decade and bring novel databases into existence. However, several long-running curated repositories for biological reference data (legacy databases) have failed to respond to the needs of their scientific communities and do not provide up-to-date services, for example, programmatic access via application programming interfaces (APIs), fine-grained and transparent data versioning and the use of persistent identifiers (PIDs). Replication, derivation and recreation of such legacy databases may address these gaps, but such activities may be hampered by the *sui generis* database right, a form of intellectual property rights (IPRs), enshrined in the EU's Database Directive.

Public biological databases are sometimes subject to central creation and central management. For example, a research institution or consortium develops or brings together related datasets for the purposes of future analysis or future use.¹ In other instances, a central institution or a research consortium operates a platform on which external contributors can integrate their own datasets for the purpose of making such data available for downstream use.² Operators of the central platforms are often responsible only for hosting the provided datasets. Other times, operators also provide value-added services to data contributors. Value-added services include curating the data to improve its technical interoperability with other available datasets or creating enriched datasets from the original input data via data analysis, data aggregation, or data visualisation tools.³ Alternatively, some public biological databases are decentralised, with a formal or informal arrangement of research institutions or citizen scientists each contributing data, cloud computing, or cloud storage resources, and analysis tools for common purposes without appointing a central data custodian.⁴

Here, we discuss the implications of the EU Database Directive on the establishment by research communities of biological data resources, including both databases and value-added services. Our legal analysis raises a number of policy considerations relevant to the functioning of public biological databases. For our purposes, public biological databases refer to collections of biological data, or

¹ O'Doherty et al. (2021).

² Poupon et al. (2017). See also Malsagova et al. (2020).

³ Bui et al. (2017). See also Olk and West (2020); Maccari et al. (2020).

⁴ Saez-Rodriguez et al. (2016); Ahsanuddin et al. (2016), p. 29; Greshake et al. (2014).

secondary outputs derived from such data, that are made available to the public using an open access, registered access, or controlled access mechanism.⁵ In Sect. 2, we discuss the relationship between the public policy justifications for the creation of intellectual property rights. In Sect. 3, we discuss the law, jurisprudence, and doctrine establishing the ambit and scope of application of the EU *sui generis* database right. In Sect. 4, we consider the potential for the *sui generis* database right to find application to public biological databases. We also consider the potential for downstream services that enable data discovery, and analysis to infringe the *sui generis* database right. Section 5 details public policy reforms that may better align the database right with contemporary approaches to open science and reduce the transaction costs inherent in negotiating downstream rights in IPR-encumbered data. In Sect. 6, we address the potential repeal of the *sui generis* database right. Section 7 outlines generalisable conclusions directed to the broader effort to establish EU information policy.

2 Intellectual Property Rights versus Open Science

2.1 Justifications For and Against Intellectual Property Rights

The design of IPRs attempts to find balance between two conflicting imperatives, first, to provide creators of intangible goods with protection from reuse of their goods absent suitable compensation and/or authorisation, and second, to enable the secondary use of existing intangible goods, especially for those that are necessary economic inputs into other productive activities.⁶ In balancing these interests, IPRs are limited in scope and duration. Upon the expiration of IPRs, the protected intangibles and associated rights of use return to the public domain. Thus, IPRs are commonly conceived of as a bargain between the creator and society, because these rights are granted to incentivise the disclosure of the intangible goods. In protecting the rights of creators in intangibles, IPRs enable producers of such goods to disclose them to third parties or to the public while retaining the ability to exclude others from using them, or to impose conditions on their use. In their absence, trade secrets protect information of commercial value that is held confidentially.⁷

IPRs are designed to address the unique character of intangible goods, which can be considered non-rivalrous, in whole or in part. The use or consumption of non-rivalrous goods by one individual does not inhibit another's use or consumption of the same goods. For example, the use of a dataset as part of an analysis does not preclude others from using the dataset afterwards.⁸ This is distinct from tangible goods whose consumption is rivalrous, in that the consumption of the good exhausts it in whole or in part and precludes its future use. Further, absent reliance on specialised technological tools, such as digital rights management (DRM)

⁵ Dyke (2020).

⁶ Contreras (2018).

⁷ Trade Secrets Directive, Art. 2(1). See also Desai (2018), p. 483.

⁸ Huarrag Guerrero (2017), pp. 83–84.

technologies, intangible goods, such as datasets, are usually non-excludable, meaning that it is usually not possible to stop a user from making a second copy and transmitting it to another recipient.

IPRs are an instrument of industrial or economic policy in that they provide a measure of excludability for a set period of time, thereby creating markets in intangible goods (IPRs can also act to condition access to a commons upon respect for its rules). Examples of IPRs include copyright protection for literary works, software, databases, and patent protection for new and useful or industrially applicable inventions.⁹ IPRs enable their holders to control access to their protected intangible goods, via licenses or assignment to third parties. The latter are forms of contracts which specify enforceable rights over intangible goods between contracting parties and largely fall within the realm of private law. Public law may provide further limitations on use of information, for example, regulatory oversight over personal data.¹⁰

Some economists argue that absent the creation of IPRs, the potential for third parties to free-ride on the efforts and investments of others will create market failures that disincentivise investment in the production of intangible goods.¹¹ In other words, these goods create positive externalities that benefit third parties who are unwilling to bear the costs of purchasing such goods. If the producers of intangibles cannot capture value through market transactions, they may opt not to produce them or to make them available. The resultant lack of a competitive market may result in suboptimal investment in the production of intangible goods.¹² Further, parties that bear the costs of first production of an intangible good are exposed to the risk that other distributors will engage in competition without bearing the fixed cost of original production of the intangible good. That is, absent IPRs, the original distributor that bears the cost of producing an intangible good is left at a competitive disadvantage relative to other distributors. Imposing costs upon third parties as a precondition to accessing intangibles, however, deprives these parties of benefits that could be realised at little or no cost, because sharing an intangible good creates additional value at no additional cost.¹³

Therefore, insofar as free markets fail to generate sufficient investment in the creation of intangibles, which may be considered “public goods”, it may be beneficial to recognise time-limited IPRs to stimulate investment in their production. Exclusive, albeit time-limited IPRs guarantee the creators who bear the cost of production a minimum market share that entitles them to sufficient profits to justify the risk of investment.¹⁴ However, “strong” IPRs that permit the unilateral and categorical exclusion of third parties from the use of intangible goods

⁹ Huarag Guerrero (2017), pp. 83–84.

¹⁰ General Data Protection Regulation.

¹¹ Hadfield (1988), pp. 16, 19. *See also* Varian (1998), p. 9.

¹² Hadfield (1988), pp. 33–35.

¹³ Hadfield (1988), pp. 22–23.

¹⁴ Hadfield (1988), pp. 16, 19.

can inhibit innovation.¹⁵ This inhibition arises from a rights-holder precluding access to intangible goods that are necessary inputs into the creation of downstream innovations, thereby maintaining market dominance. It also arises from the transaction costs entailed in negotiating access to all of the intangible goods required for an anticipated purpose. For example, in the context of data-intensive biological research, IPRs over aspects of data can limit the value derived from generating and combining datasets of disparate origin, in part, due to legal uncertainty. Attempts to train machine learning (ML) algorithms might require access to a large number of datasets that are each potentially subject to copyright, moral rights, or *sui generis* database rights. Though the creators of such datasets may have no interest in precluding third parties from using them, the default operation of law may invest such creators with legal rights to exclude third parties from using their protected datasets. Third parties, therefore, lack certainty about whether their intended use of a dataset will infringe IPRs. This encourages conservatism in the use of available intangibles, due to real or imagined risks of legal non-compliance.¹⁶

Gauging the amount of protection to confer via IPRs is difficult. Over-protection might stifle the pro-social use of existing intangibles, whilst under-protection might inhibit investment in their initial generation. Studies demonstrate the complex relationship between IPRs and domestic economic growth or contribution to innovation. IPRs best contribute to economic development when they are time-limited,¹⁷ and their economic benefit is contingent on other characteristics, such as access to a stable enforcement infrastructure. IPRs best benefit developed economies that have a strong innovation sector but can prove a detriment to developing economies without significant domestic IPR holdings. The latter benefit more from liberal access to intangible goods developed elsewhere.¹⁸

For these reasons, legislators and courts have adopted numerous measures to nuance the relationship between intellectual property protection, innovation, and the

¹⁵ Lipton (2009), pp. 635–636, 638, 641–642. Lipton contends that the creation of strong IPRs in datasets and other online content during the emergence of the internet anticipated strong rights in protected content matter to stimulate investment in the production thereof, and anticipated individual natural persons and legal entities, or small groups, as the producers of datasets and online content. This proved detrimental to the later flourishing of community-led content-creation initiatives that benefit from the reutilisation and modification of data and content contributed to the internet, which represents a greater portion of online data generation and content generation than do the single-person or single-entity initiatives to create content that strong and individuated IPRs incentivise.

¹⁶ Eisenberg (2008), pp. 1075–1088. Eisenberg posits that empirical evidence demonstrates that intellectual property rights have given rise to an “anticommons” in biomedical research inputs, such as patents, and especially contractual or intellectual property rights in datasets. The principal challenges are the existence of exclusive property rights in datasets, and the high transaction costs inherent in negotiating access to biomedical data. Eisenberg does nuance her position in concluding that for public materials (e.g. datasets) for which the burden of enforcement rests on the right-holder, rather than the prospective user, the precluding effect of such rights is more limited. More limited, that is, compared to circumstances where the prospective research must lodge an application for access to the protected research materials. See also Rai (2017), p. 1646 (albeit Rai’s discussion centralises patents rather than other, principal, intellectual property rights in datasets and databases).

¹⁷ Posner (2005), pp. 58–62.

¹⁸ Cunha Neves et al. (2021); Drahos and Braithwaite (2002), pp. 3, 11–13.

prosocial use of existing intangible goods. Legislatures may create new IPRs or fine tune existing ones through statutes or regulations. Courts contribute by interpreting the scope of IPRs, for example, allowable subject matter for IPRs and tests for IPR criteria, or in devising exceptions to the application of IPRs. In so doing, courts avoid restricting the downstream use of infrastructural intangible goods.¹⁹ These act as necessary inputs into downstream production processes, and limitations in access or use could preclude economic activity and innovation. Courts nuance the application of IPRs for reasons of public policy, but also increase the overall complexity of the applicable legal rules, especially in supranational legal systems, such as that of the EU. For example, the national courts of individual EU Member States can create legal fragmentation through heterogeneous interpretation of supposedly harmonised inter-State intellectual property law regimes. In partial response to the difficulties inherent in administering IPRs and determining their boundaries, and in recognition of the increasing role of non-market institutions (e.g. academia, crowdsourcing) in producing intangibles, open science initiatives are coming to be seen as a privileged vehicle for incentivising the production of valuable intangibles that are minimally encumbered by private rights.

2.2 Justifications For and Against Open Data in Support of Open Science

Open science is an umbrella term that captures an ongoing shift in the policy considerations that determine how the outputs of scientific research are to be disseminated, and which categories of actors can participate in the scientific enterprise. Though there is no consensus definition of open science, the concept has been recognised in numerous international policy documents, including the recent UNESCO Recommendation on Open Science²⁰ and multiple OECD instruments.²¹

The policy justifications for open science are numerous and contested. Central themes in the open science literature include the following: (1) democratising science to enable participation and comprehension of research results by members of the public; (2) decreasing research costs by enabling access to inputs such as software, data, and publications, as well as infrastructure such as computational and analytic resources; (3) increasing the accountability of scientific researchers in making data, methodologies, and results more accessible to enable replication and validation of findings.²²

Open science initiatives acknowledge public participation in defining research agendas and performing research (e.g. citizen science), and produce mechanisms that foster research reproducibility and scientific integrity.²³ However, they also

¹⁹ Lee (2008).

²⁰ United Nations Economic, Social, and Cultural Organisation (UNESCO) (2021).

²¹ Organisation for Economic Cooperation and Development (OECD) 2021 “Recommendation of the Council concerning Access to Research Data from Public Funding”; Organisation for Economic Cooperation and Development (OECD) 2021 “Open Science: Enabling Data Discovery in the Digital Age.”

²² Fecher and Friesike (2014).

²³ Ross et al. (2012).

reflect pragmatic changes in science policy that flow from technological advances. As the costs of storing and transferring data decline, and as data are increasingly generated as a residual output of other activities (e.g. clinical care or public-sector service provision), the potential to repurpose existing data to drive novel research efforts at low cost becomes easier. Therefore, the trend toward privileging open science activities also reflects a response to fundamental changes in the cost-benefit proposition of repurposing existing data to enable further downstream research and of the heightened potential to coordinate the activities of scientific researchers and the public toward cooperative efforts using decentralised platforms.²⁴

Open science raises policy tensions. The determination of the best combination of incentives and institutions to generate intangibles and translate them into downstream outputs is often context-specific and indeterminate. For example, if generating data is cost-intensive and difficult to coordinate through non-market institutions, it might be preferable to use property rights in intangibles to stimulate production. In contrast, in those instances where researchers, public sector institutions, or members of the public have strong incentives to generate primary data or downstream intangibles absent market-based incentives, it might prove more effective in generating downstream research outputs to require the open dissemination of data and to limit or exclude the application of IPRs. However, placing limitations on the potential to commercialise downstream outputs resulting from the use of data and other intangibles must be done with great caution; limiting the prospect of commercialising or using IPRs to protect downstream innovations such as drugs, therapies, or software tools could have a chilling effect on their production.

In practice, open science, and open access to data more generally, has been championed through a number of different approaches. Government and public-sector approaches include requirements to release biomedical research data and public sector data to the public. National funding agencies impose on researchers open science terms that apply to data generation and other intangible outputs. Some research communities and consortia voluntarily enable the open use of their data and other outputs. Mechanisms include open publication, licenses, or dedications to the public domain to enable downstream access.

3 EU Laws Relevant to Database Rights and Open Science

3.1 The *Sui Generis* Database Right

The *sui generis* database right is a novel IPR first implemented in EU law, and subsequently adopted in three other jurisdictions.²⁵ This right protects the contents of a database from the unauthorised “extraction and/or or re-utilisation of the whole

²⁴ Benkler (2007).

²⁵ Doldirina et al. (2018). According to the above authors, these include: Mexico, Russia, and South Korea. The protection recognised in Mexico, however, is a form of copyright that does include an originality standard, rather than a *sui generis* database right in the European Union’s sense of the term. See World Intellectual Property Organisation (WIPO) (2002).

or of a substantial part.”²⁶ This right is applicable to databases if a “substantial investment”²⁷ is made in their compilation. The Database Directive elaborates on what comprises the extraction and/or reutilisation of a “substantial portion” the database. The “substantial part” criterion is assessed from both a quantitative *and* a qualitative perspective.²⁸ Consequently, the database right could be qualitatively infringed, even if the portion extracted or reutilised does not reflect a large portion of the overall contents of the database.²⁹

The right is accorded to the “maker” of the database, but “maker” is not defined and is understood to include both natural persons and legal persons (e.g. corporations) who are “nationals of a Member State or who have their habitual residence in the territory of the community.”³⁰ In contrast to copyright, which protects creative expression, the right protects the database if a “substantial investment”³¹ has been made in its compilation. Natural and legal persons can benefit from *sui generis* database protection, whilst subcontractors cannot benefit from such right.³²

This right is one of the strongest forms of IPR created by legislatures, because its term is indefinite and there are only limited exceptions to its application. The right expires after a 15-year period.³³ However, because the right is re-actualised each time a database is modified³⁴ (i.e. the 15-year period restarts), the duration of the *sui generis* database right is functionally unlimited.³⁵ Unlike its closest IPR relative, copyright, it lacks fair-use exceptions and other similar limitations.

EU Member States can implement limited exceptions in national legislation. First, an exception can be applied to extraction that is performed for the purpose of non-commercial teaching or non-commercial scientific research activities.³⁶ Second, an exception can be made for the purposes of private use of non-electronic

²⁶ Database Directive, Art. 7(1).

²⁷ Database Directive, Art. 7(1).

²⁸ Database Directive, Art. 7(1).

²⁹ Database Directive, Art. 7(1).

³⁰ Database Directive, Art. 11(1).

³¹ Database Directive, Art. 11(1).

³² Database Directive, Recital 41. Recital 41 of the Database Directive makes explicit that subcontractors cannot benefit from the protection of the *sui generis* database right (i.e. the right does not arise in their favor, by operation of law). Much of the legal academic literature has inferred that employees cannot benefit from the protection of *sui generis* database rights, either. This is usually justified analogously from the preclusion imposed on subcontractors from benefiting from the *sui generis* database right, or through the interpretation of the investment criterion (such investment being a legal precondition to the right arising in favor of a prospective rightsholder). Certain authors instead consider that the determination of whether employees can benefit from the *sui generis* database right is left to the determination of local Member State law. See also European Commission, Study in support of the evaluation of Directive 96/9/EC on the legal protection of databases (2018).

³³ Database Directive, Art. 10(1).

³⁴ Database Directive, Art. 10(3). The modifications effected must themselves constitute a “substantial new investment.” However, the Database Directive explicitly contends that multiple successive minor amendments can additively constitute such a “substantial new investment.”

³⁵ Database Directive, Art. 10(3).

³⁶ Database Directive, Art. 9(b).

databases.³⁷ Third, Member States can introduce an exception for purposes of “public security or an administrative or judicial procedure.”³⁸

Text and data mining (TDM) activities do not infringe the extraction prong of the *sui generis* database right,³⁹ unless the holder of the *sui generis* database right makes an express preclusion of such activities.⁴⁰ TDM performed for scientific research purposes does not infringe the extraction prong of the *sui generis* database right, *regardless* of whether or not the holder of the *sui generis* database right expressly precludes such activities.⁴¹

Finally, the Database Directive limits the extent to which the holders of *sui generis* database rights can restrict the actions of lawful users of protected databases.⁴² Rights-holders cannot contractually assert rights in databases that go beyond the protection provided by the *sui generis* database right. Specifically, the rights-holder cannot, contractually or otherwise, preclude lawful users from extracting or reutilising insubstantial portions⁴³ of protected databases.⁴⁴ One of the critical challenges inherent in the interpretation of this provision is that “lawful user” is not a defined term in the Database Directive, and Member States have diverged in their definitions in their local implementations of the Database Directive.⁴⁵

³⁷ Database Directive, Art. 9(a).

³⁸ Database Directive, Art. 9(c).

³⁹ Digital Single Market Directive. The definition of “text and data mining” is as follows: “any automated analytical technique aimed at analysing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations.”

⁴⁰ Digital Single Market Directive, Arts. 4(1), 4(3).

⁴¹ Digital Single Market Directive, Arts. 3(1), 3(2), 4(3), 4(4). This right further incorporates a right to retain the concerned databases for the purpose of performing the verification of research results.

⁴² Database Directive, Art. 8(1).

⁴³ Database Directive, Art. 7(1). The Database Directive does not provide meaningful guidance concerning the distinction between substantial and insubstantial portions of a database, apart from the stipulation that a portion can be substantial in a qualitative or quantitative fashion. This creates challenges in determining whether the “extraction” or “re-utilisation” of part of a database will concern an insubstantial portion, which implies that it is permitted through the operation of law, or whether it will concern a qualitatively or quantitatively substantial portion, which breaches the *sui generis* database right. This criterion can prove especially pernicious to navigate as the portions of a database deemed “qualitatively” substantial are presumably those that are of greatest interest to secondary users.

⁴⁴ Database Directive, Art. 7(1).

⁴⁵ European Commission (2018), pp. 11–12. The 2018 European Commission study establishes the situation as follows. Most implementations of the Database Directive do not incorporate a definition of “lawful user.” Of those that define the term, there is disagreement as to two principal elements. First, the categories of users that can be considered “lawful users”, and second, the ambit of acts that fall within the umbrella of “lawful use.” Select jurisdictions (*e.g.* France and Malta) restrict such protection to parties having a contractual right to access the concerned database; others use more generous language such as those having an “authorisation” (*e.g.* Austria and Hungary) or a “right to use” (*e.g.* Denmark, Finland, and Sweden). The law of Ireland and United Kingdom referred to those uses provided for “under license or otherwise.” German legislation has implemented a broad definition of the term “lawful user”, citing a large number of acceptable circumstances, such as with the consent of the rights-holder, according to contract, or having purchase a copy of such database, irrespective of the conditions of purchase. The European Commission study also notes the following distinct approaches to defining the categories of

3.2 *Sui Generis* Database Right: Historical and Policy Context

While the *sui generis* database right provides considerable protection, the scope of its application has been limited by EU courts.⁴⁶ Understanding this limitation from a policy perspective requires consideration of the historical origins of the database right. In the late 1980s, EU policymakers convened to reform EU intellectual property law,⁴⁷ for the purpose of enhancing EU competitiveness in the emergent digital technologies sector. An objective of the reforms was to incentivise the production of data to match that of the United States.⁴⁸ In crafting the EU Database Directive, policymakers first harmonised the copyright protections afforded to databases, while respecting the distinct copyright traditions of EU Member States.⁴⁹ This approach attempted to harmonise the intellectual property protection afforded to databases throughout the EU, while respecting the distinct intellectual property traditions of continental Europe, the United Kingdom, and those of select other (mostly Nordic) EU Member States.⁵⁰

Historically, continental Member States required a compilation, such as a database, to demonstrate “originality” to be considered eligible for copyright protection. Conversely, the United Kingdom, adopted a more liberal “sweat of the brow” doctrine, which provides copyright protection if effort, labor, or other resources are expended in the creation of a database. The Database Directive adopted the continental approach to determine whether subject matter was eligible for copyright protection throughout the EU⁵¹ This approach lowered the legal standard required to obtain copyright protection in a select number of EU jurisdictions, such as Germany, but raised it in the United Kingdom.⁵²

To offset concerns that the ensuing protection afforded by copyright to databases was too narrow, EU policymakers created the *sui generis* database right.⁵³ This right strengthened IPR protections available to database producers⁵⁴ by protecting databases where an “investment”, considered with respect to capital, human labor, or technological resources was made in their production.⁵⁵ It applies legal protection for compilation of pre-existing data that was historically recognised in the Nordic

Footnote 45 continued

uses which this right protects, including “use”, “normal use”, “intended purpose”, or as established through contract.

⁴⁶ C-203/02 *The British Horseracing Board Ltd and Others v. William Hill Organization Ltd*. ECLI:EU:C:2004:695 (hereinafter *British Horseracing Board*). See also Case C-762/19, *CV-Online Latvia SIA v. Melons SIA* ECLI:EU:C:2021.

⁴⁷ Hugenholtz (2000).

⁴⁸ Hugenholtz (2016), pp. 205–206, 221–22. See also Database Directive, preamble at para. 11.

⁴⁹ Hugenholtz (2016), pp. 3–4.

⁵⁰ Hugenholtz (2016), pp. 6–7.

⁵¹ Hugenholtz (2016), pp. 6–7.

⁵² Hugenholtz (2016), pp. 6–7.

⁵³ Hugenholtz (2000), pp. 3–4.

⁵⁴ Hugenholtz (2000), pp. 3–4.

⁵⁵ Database Directive, preamble at para. 40.

EU Member states⁵⁶ and can be understood as applying the United Kingdom's "sweat of the brow" approach, despite the rejection of this standard for copyright.⁵⁷

The EU initially advocated for the database right to be implemented in other jurisdictions, and for its transliteration into an international standard of intellectual property protection.⁵⁸ However, numerous commentators have expressed significant reservations about the right's effectiveness in stimulating investment in the production of databases since its implementation.⁵⁹ Indeed, the European Commission has itself joined in the critique of the Database Directive and its unintended deleterious effects.

In December 2005, the Commission released its first evaluation of the Database Directive since its coming into force. The evaluation considered whether the Database Directive has achieved its stated policy objectives, and considered whether the *sui generis* database right has stifled economic activity by precluding competition.⁶⁰ The European Commission concluded that there was no empirical evidence to suggest that the *sui generis* database right had generated increased investment in database production and that there had been no demonstrable increase in database production in the EU, either before or after the implementation of the *sui generis* right.⁶¹ Instead, research stakeholders interviewed for the report suggested that the database right had led to the strengthening of monopolies on database production and dissemination amongst the largest market participants.⁶² The result is not surprising based on IPR economics literature, much of which concludes that the creation of strong IPRs of long duration entrenches the economic advantages of established producers or aggregators of intangible goods, rather than bolstering the competitiveness of new market entrants.

The European Commission concluded that harmonisation efforts had not succeeded because of ambiguous language used to define the scope and subject matter of the *sui generis* database right.⁶³ These ambiguities are more pronounced for the *sui generis* database right than for copyright, which has been streamlined

⁵⁶ Hugenholtz (2016). Specifically, these are: Denmark, Finland, Iceland, Norway and Sweden.

⁵⁷ Beunen (2007), p. 140. Beunen cites Speyart 1996 in support of this conclusion.

⁵⁸ Hugenholtz (2016), pp. 209–210.

⁵⁹ Cardinale (2006–2007). Cardinale describes the response of regulators, the legislature, and academic commentators in the United States to the EU Database Directive, including numerous initiatives to introduce an analogue to the EU *sui generis* database right in United States legislation. In the United States, efforts to implement a *sui generis* database right have often been bipartisan in nature, albeit obtaining much more support from Republican legislators than from Democrats (p. 164). The principal reservations against the implementation of the *sui generis* database right include differing policy positions on the strength and ambit that should be accorded to the protection, the fear of stifling academic enterprise and scientific research, and doubts regarding the efficacy and benefits to be derived from this legal innovation (p. 168).

⁶⁰ European Commission (2005), p. 3.

⁶¹ European Commission (2005), pp. 5–6, 20.

⁶² European Commission (2005), pp. 5–6, 20.

⁶³ European Commission (2005), pp. 23–24.

through a rich history of national legislation implementation and jurisprudential interpretation. The database right left national courts to adopt novel and often conflicting interpretations.⁶⁴ Divergence in interpretation further arose from the implementation of the Directive in the national law of each Member State.⁶⁵ In addition, the European Commission found that the breadth of the *sui generis* database right almost amounted to a “pure” property right in raw data.⁶⁶

In contrast, consulted members of the EU database industry overwhelmingly felt that the database right protected them well, and that EU harmonisation of rights in databases was of crucial importance.⁶⁷ This view did not extend to representatives of scientific research bodies, libraries, and academic institutions, who decried the narrow exception for scientific research and its restriction to non-commercial research use.⁶⁸

⁶⁴ Stamatoudi (1997), pp. 443–445. In 1997, Stamatoudi correctly anticipated that EU courts would adopt conflicting interpretations of the scoping criteria *sui generis* database right (*i.e.* the ambit of the materials captured), due to numerous definitions in the directive that are open to divergent interpretation. Stamatoudi determined that the language used to define “database” created ambiguities as to its scope of application. Database is defined as a “collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means.” (Database Directive, Art. 1). Stamatoudi contends that the terms “collection”, “independent works”, “arranged in a systematic or methodical way”, and “individually accessible” each lend themselves to significant interpretive ambiguities.

⁶⁵ Stamatoudi (1997), pp. 446–447; Hugenholtz (2000), pp. 15–19; Hugenholtz (2016), p. 221. Hugenholtz describes how numerous Member States initially integrated to their national legislative implementations of the *sui generis* database right – or copyright – the continued subsistence of their non-harmonised, pre-Directive approaches to copyright protection, frustrating the Database Directive’s attempt at implementing a consistent standard of IP protection for databases throughout the Union. However, most EU Member States have in later years harmonised their intellectual property regimes to the standards established in the Database Directive. In contrast to Hugenholtz, a number of other authors have identified considerable remaining incompatibilities in the national implementations of both the *sui generis* database right and copyright portions of the Database Directive. Such authors also note that other elements of Member State national law (*e.g.* rules governing property rights and choice of law) can lead to differing outcomes in the interpretation of the local *sui generis* database right and copyright applicable to databases. See Koščík and Myška (2017). Koščík and Myška describe a number of potential legal uncertainties that can arise from the distinctions in local Member State law governing *sui generis* database rights, including the general civil law of Member States, and the local statutory implementation of the Database Directive. These are: differences in the language used to define “substantial investment” in national implementations of the *sui generis* database right; in the legal test used to determine which national intellectual property law should govern a legal dispute (*i.e.* whether the law of the jurisdiction in which the dispute arises, or the law of the jurisdiction in which the protected work was produced, should supersede); dissimilarities in the applicable preconditions to the recognition of a property right in the Member State’s general property law rules; incompatible approaches to the attribution of “joint authorship” in domestic intellectual property law; and discrepancies in the potential to use contracts to transfer, assign, license, or waive intellectual property rights.

⁶⁶ European Commission (2005), p. 24.

⁶⁷ European Commission (2005), p. 20.

⁶⁸ European Commission (2005), p. 21.

The 2005 evaluation⁶⁹ was followed by a rigorous, empirical, 2018 study in support of a second evaluation,⁷⁰ which assessed the economic and legal effects of the Database Directive. The 2018 study focused on the perspectives of industry participants, such as “database users”, “database user-producers”, and “database producers”, on the economic and legal effects of the Database Directive.⁷¹ Representatives from research institutions, academic institutions, and legal experts also contributed to the study.⁷²

The authors of the 2018 study considered the effectiveness of the *sui generis* database right and made reform recommendations. They concluded, first, that the economic value of data is increasingly realised through the combination, analysis, and modification of existing datasets. Thus, granting database creators exclusive rights in databases can deter productive activities.⁷³ Second, the exceptions to the *sui generis* database right are more limited than for copyright,⁷⁴ which creates barriers to the secondary use of existing databases. Third, the content and ambit of the *sui generis* database right remain unclear, both due to drafting ambiguities and divergent statutory implementation and jurisprudential interpretation of the Directive in each Member State.⁷⁵ Fourth, the Court of Justice of the European Union (CJEU)’s holding in *Ryanair Ltd v. PR Aviation VB* (discussed below) enabled database producers to hold their IPR-ineligible databases to a higher standard of protection than if their databases were eligible for *sui generis* database right or copyright protections, through contractual restrictions on their use.⁷⁶ Finally, the database right coexists uneasily with later EU initiatives to foster the creation of open data repositories and requirements for freely accessible public sector information (PSI).

Proposed policy reforms strike a balance between maintaining the benefits of the *sui generis* database right and limiting its more detrimental features. They include:

⁶⁹ Cardinale (2006–2007), pp. 173–174. Cardinale and other commentators have expressed significant concerns regarding the methodology of the first European Commission evaluation of the *sui generis* database right. Difficulties include high reliance on the opinions of stakeholders that benefit from the preservation of the *sui generis* database right (e.g. database producers), and the under-reliance on empirical data concerning database production from EU and non-EU jurisdictions, and a relative unwillingness to seriously consider the repeal or reworking of the *sui generis* database right despite significant empirical evidence that suggests this could be the best course of action. These critiques also appear salient, albeit to a lesser degree, to the second evaluation of the *sui generis* database right, discussed below.

⁷⁰ The methodologies used are qualitative in nature, and include: online surveys, semi-structured interviews and an in-person workshops targeted at concerned stakeholders, and public consultation.

⁷¹ European Commission (2018), pp. 1–3.

⁷² European Commission (2018), pp. 1–3.

⁷³ European Commission (2018), pp. 26–27.

⁷⁴ European Commission (2018), pp. 18–19.

⁷⁵ European Commission (2018), pp. 7–10. Divergent elements include: the threshold of substantial investment required to give rise to the *sui generis* database right (jurisprudence and statutory), unique implementations of the exceptions to the *sui generis* database right in each Member State, including the addition of novel exceptions not articulated in the Database Directive in select Member States (statutory), and the definition of a “substantial” part of a database (i.e. what threshold of reutilisation or extraction triggers a violation of the *database right*) (jurisprudential).

⁷⁶ European Commission (2018), pp. 15–16.

(1) implementing a registration requirement or other legal formalities as a precondition to the application of the *sui generis* database right, rather than having the rights apply through default operation of the law;⁷⁷ (2) expanding exceptions to the *sui generis* database right, such as those applied to copyright in the Information Society Directive;⁷⁸ (3) introducing a compulsory licensing scheme for sole-source databases;⁷⁹ and (4) implementing legislative reforms to overturn the CJEU's *Ryanair* decision.⁸⁰

3.3 *Sui Generis* Database Right: Jurisprudential Interpretation and Other Limiting Factors

Having addressed the substantive contents and the historical evolution of the *sui generis* database right, we now discuss the judicial interpretation of that right by the CJEU and the national courts of EU Member States. This jurisprudence is consistent with the conclusion that the *sui generis* database right serves to protect investment in producing a database more than the database itself.

In *The British Horseracing Board and Others*, the CJEU interpreted both the breadth of the database right and the circumstances that give rise to it.⁸¹ The CJEU interpreted the substantive contents of the right expansively, concluding that the “quantitative” prong of the statutory test protects the extraction or the reutilisation of a large proportion of the database’s contents.⁸² However, the right also protects a small proportion of the contents of a database if the concerned proportion represents a significant amount of the effort or other resources utilised in the creation of the database.⁸³ Further, the court established that the *sui generis* database right is not exhausted if the maker of the database renders the database public, either directly or by allowing a third party to make the database public. In other words, the database right that limits extraction or reutilisation continues to apply to databases that have been made public.⁸⁴ Merely consulting a database, however, does not constitute extraction or reutilisation, and therefore does not violate the *sui generis* database

⁷⁷ European Commission (2018), pp. 32, 65.

⁷⁸ European Commission (2018), p. 19. See also Information Society Directive, Art. 5.

⁷⁹ European Commission (2018), pp. 34–37, 65–66, 141. The authors of the study further note that the original draft of the Database Directive incorporated a compulsory licensing scheme directed to sole-source databases; however, this was omitted from the final draft of the Database Directive due to disagreement on the part of select Member States. The authors of the study interpret these intended compulsory licensing provisions as a counterbalancing element against the strength and duration of the *sui generis* database right. The removal thereof might therefore be responsible for producing a lopsided situation in which the protections that the *sui generis* database right affords unduly privilege the interests of the protected database maker over the public’s right to access and use databases.

⁸⁰ European Commission (2018).

⁸¹ *The British Horseracing Board Ltd and Others v. William Hill Organization Ltd*. ECLI:EU:C:2004:695.

⁸² Database Directive, Art. 7(1).

⁸³ *British Horseracing Board* at para. 71.

⁸⁴ *British Horseracing Board* at paras. 55–58.

right.⁸⁵ Repeated extraction or reutilisation of an insubstantial portion of a database can cumulatively lead to the breach of the *sui generis* database right.⁸⁶

Despite its liberal interpretation of the breadth of the *sui generis* database right, the CJEU in *British Horseracing Board* went on to restrict its application. The CJEU held that the “substantial investment” protected by the database right is the investment in compiling the database, *not* the investment in creating or producing the data contained in the database.⁸⁷ In further jurisprudence, the CJEU confirmed that investment in producing the contents of a database are of no relevance to satisfying the prerequisite of “substantial investment” for a *sui generis* database right to arise.⁸⁸ The CJEU stipulates that the creator of data can in some instances claim a *sui generis* database right in a database assembled from data that it has generated, but that the efforts made to generate such data do not contribute to establishing a legal entitlement to the *sui generis* database right.⁸⁹ The CJEU justifies its restrictive interpretation based on its reading of Recital 39 of the Database Directive,⁹⁰ which states that the purpose of the right is to incentivise the organisation of data into datasets and the maintenance of those datasets, rather than to foster the generation of data.⁹¹

Commentators agree that the Database Directive applies to “collected” databases but not to “created” databases, and that this interpretation is consistent with the Database Directive’s objective to protect investment in the curation of new databases. Our own interpretation is that the “creation” of data is used as a “proxy” for the underlying policy issue that the courts are trying to address: denying protection to sole-source databases, to foster competition in the market for information and related services. That is, database-production efforts that entail the primary creation of data have a high likelihood of producing “sole-source” databases.⁹² It is easier to verify whether a database is “created” than it is to

⁸⁵ *British Horseracing Board* at paras. 54–58.

⁸⁶ Database Directive, Art. 7(5). University of Amsterdam (2006) citing: *Spot (cinebel.be) v. Canal Numédia (allocine.be)* District Court, Brussels, 18 January 2002.

⁸⁷ *British Horseracing Board* at paras. 31–33, 79.

⁸⁸ Case C-46/02 *Fixtures Marketing Ltd v. Oy Veikkaus Ab* ECLI:EU:C:2004:694 at paras. 38, 45–47. Case C-444/02 *Fixtures Marketing Fixtures Marketing Ltd v. Organismos prognostikon agonon podosfairou AE (OPAP)* ECLI:EU:C:2004:697 at para. 43. C-338/02 *Fixtures Marketing Ltd v. Svenska Spel AB* ECLI:EU:C:2004:696 at paras. 21, 23–24, 27, 29. (referred to as *Oy Veikkaus*, *OPAP*, and *Svenska Spel*, respectively). The CJEU stipulates that “resources used, with a view to ensuring the reliability of the information contained in that database, to monitor the accuracy of the materials collected when the database was created and during its operation” are of relevance, as are “the resources used for the purpose of giving the database its function of processing information, that is to say those used for the systematic or methodical arrangement of the materials contained in that database and the organisation of their individual accessibility.”

⁸⁹ *Oy Veikkaus* at paras. 33–36, 38–40. *OPAP* at paras. 39–40, 45–46. *Sevenska Spel* at paras. 25, 29–30.

⁹⁰ *Oy Veikkaus* at paras. 33–36, 38–40. *OPAP* at paras. 39–40, 45–46. *Sevenska Spel* at paras. 25, 29–30. The CJEU notes that Recital 39 refers to: the practice of “Obtaining and collection of the content.” The CJEU and the Advocate General consider this wording to exclude investment in the generation of data.

⁹¹ *Oy Veikkaus* at paras. 33–36, 38–40. *OPAP* at paras. 39–40, 45–46. *Sevenska Spel* at paras. 25, 29–30.

⁹² The initial conception of the Database Directive created exceptional rules applicable to sole-source datasets (*i.e.* compulsory licensing requirements requiring rights-holders to enable third parties to make commercial use) to preclude the Directive from producing monopolies in raw data. The Directive as

perform a deeper analysis of whether it is a sole-source database. Therefore, the creation/collection test is used to distinguish non-protected databases from protected databases.

Further, databases composed of data generated without an independent investment through the regular business activities or functions of an institution or an organisation do not require additional legal protection (these are referred to as spin-off databases). No public policy justification exists to protect such data, because these data are generated regardless of the incentive value that the *sui generis* database right might provide.⁹³ Courts have therefore denied legal protection to spin-off databases through the requirement of an independent investment to be made in the compilation of a database.

Thus, the scope of protection afforded by the database right is broad, once such a right is confirmed to exist. Nonetheless, the application of the database right is curtailed by the narrow interpretation of eligible “investments” in producing databases, which excludes investment in the generation or production of data, and the incidental compilation of data arising from other activities. The principal policy justification for this narrow⁹⁴ reading of the right is to limit the potential for data producers to assert database rights on databases comprised of sole-source data and on “spin-off” databases. Precluding the assertion of rights in “created” databases prevents the creators of valuable data from engaging in anticompetitive practices related to data, of which such created databases are often the exclusive source.⁹⁵ Precluding the assertion of rights in spin-off datasets ensures that such datasets – in which no investment was made – are not subject to the *sui generis* database right, as this would preclude downstream use of such data without the restriction serving to reward an investment in database production.

In its 2021 decision, *CV-Online Latvia SIA v. Melons (Latvia)*, the CJEU further narrowed the scope of application of the *sui generis* database right, through a restrictive interpretation of the conditions causing the breach of the database right. The defendant in *Latvia* reutilised a significant portion of a protected database. According to prior jurisprudence, this would have constituted an unqualified infringement of the *sui generis* database right. The court instead determined that the

Footnote 92 continued

adopted did not include such provisions directed to sole-source databases. See e.g. Proposal for a Council Directive on the legal protection of databases, COM(92) 24 final – SYN 393, (Brussels, May 13, 1992); Explanatory Memorandum. See also Borghi and Karapapa (2015), citing the foregoing document.

⁹³ Beunen (2007), p. 113.

⁹⁴ Beunen (2007), p. 133. Beunen contends that the policy choice to interpret the investment in “obtaining” data, and the distinction between “created” and “collected” data purposively to combat informational monopolies is counterintuitive from a policymaking perspective. According to Beunen, the first considerable limitation is that because information monopolies are combated through indirect proxies, there remains the potential for them to arise whilst still meeting the formal preconditions for *sui generis* database protection to be applicable. The second considerable limitation is that the reading-down of the *sui generis* database right can only increase the openness of databases that are made public. It cannot preclude monopolies from arising in data that is not made public, which would be better accomplished through an explicit power for courts or regulators to compel the release of valuable raw data.

⁹⁵ Myska and Harasta (2016), p. 179.

reutilisation of a significant portion of a protected database infringes the *sui generis* database right only if such reutilisation risks preventing the rights-holder from benefiting from the fruits of their original investment, enabling cost-recovery and financial profits being the *raison d'être* of the *sui generis* database right.⁹⁶ It also held that the existence of a risk to the investment was not sufficient to constitute a breach – where such risk was identified, it was required to consider whether the impugned act furthers an objective of the Database Directive that justifies the perpetuation of such risk, such as making information available for downstream use, or creating valuable downstream products.⁹⁷ The presence or absence of a risk posed to a protected investment nonetheless is, since *Latvia*, the principal criterion used to evaluate whether or not the *sui generis* database right has been breached.⁹⁸

In *Ryanair Ltd v. PR Aviation VB (Ryanair)*, the court considered the legal effects of the Database Directive on databases that are not protected by either the database right or copyright.⁹⁹ The Database Directive protects lawful users of databases against the assertion by rights-holders of restrictions that exceed those contemplated in the Database Directive.¹⁰⁰ The court in *Ryanair* therefore considered whether this “right” favoring database users was applicable to databases if neither copyright nor the database right were applicable.¹⁰¹ The court answered in the negative.¹⁰² Counterintuitively, the effect of this decision is to enable producers of databases that the Database Directive does not protect to impose even stricter protections via contract.¹⁰³ Here it is important to recollect that rights-holders cannot contractually

⁹⁶ Case C-762/19, *CV-Online Latvia SIA v. Melons SIA* ECLI:EU:C:2021:434 at paras. 41–46. See also Declaye and Husovec (2021), pp. 5–7.

⁹⁷ *Latvia* at paras. 41–42.

⁹⁸ *Latvia* at para. 44.

⁹⁹ Case C-30/14 *Ryanair Ltd v. PR Aviation BV* ECLI:EU:C:2015: 10 (referred to as *Ryanair*).

¹⁰⁰ Database Directive, Art. 8.

¹⁰¹ *Ryanair* at paras. 34–40.

¹⁰² *Ryanair* at paras. 34–40. This legal outcome is potentially controversial, as recognition of IPRs in databases to which the Database Directive applies could have been interpreted as being *independent* from the Database Directive’s protection of the right to reutilise or extract insubstantial portions of all databases that meet the definition of a database. This alternate interpretation would have established a general right to perform insubstantial acts of extraction and reutilisation on all databases, that could not be overwritten in reliance on contracts or other mechanisms. If this interpretation had been adopted, the rights of individuals to use databases in the EU would be subject to standardised minimum guarantees, and the IPRs in the Database Directive would create an entitlement to protections that benefit qualifying database makers (*i.e.* those that invest or that demonstrate creativity). Instead, the CJEU in *Ryanair* interpreted the categorical legal entitlement of third parties to extract and reutilise insubstantial portions of a database as being *contingent on* the database being IPR-protected. This is counterintuitive from a public policy standpoint. In essence, it transforms the *sui generis* database right and database copyright from strict rewards for exhibiting creativity or investment, to a mixed bag of legal benefits and detriment. Protection according to the Database Directive provides beneficiaries categorical protections against most acts of substantial extraction or reutilisation, but also restricts their otherwise unfettered contractual right to preclude co-contractants from performing insubstantial acts of extraction or reutilisation. The result is that database producers might prefer *not* to benefit from copyright or the *sui generis* database right, in order to maintain the right to impose strong contractual protections on contractual users of their databases.

¹⁰³ Myska and Harasta (2016); Bottis (2015), pp. 173–174; Borghi and Karapapa (2015) at s. “When user freedom flies away – *Ryanair v. P.R. Aviation*.”

assert rights in databases that go beyond the protection provided by the *sui generis* database right.

3.4 The Open Data Directive

In contrast to the Database Directive, the Open Data Directive encourages public sector bodies to openly share data. The Open Data Directive entered into force in 2019, amending and replacing the Public Sector Information Directive (PSI Directive) that was adopted in 2003 and first updated in 2013.¹⁰⁴ These successive Directives established baseline rules that must be integrated to the access-to-information laws of EU Member States, whilst leaving to Member States the discretion to determine which documents and information will be subject to such regimes.¹⁰⁵

The PSI Directive implemented a minimal set of rules enabling the reuse of documents that public sector bodies made available to the public and authorised the reuse of. These rules did not compel public sector bodies to make their documents available to the public, nor to authorise their reuse. However, the PSI Directive applied when public sector bodies voluntarily chose to make their documents public for the purpose of reuse (or national law required them to do so). It created obligations that determined the conditions according to which such data were to be shared.¹⁰⁶ The PSI Directive mandated that public sector bodies provide access to their documents for commercial and non-commercial purposes. Such bodies were required to respond to requests for the use of documents expediently, limit charges levied to a “reasonable return on investment”,¹⁰⁷ subject access to minimally restrictive licenses,¹⁰⁸ and provide all market participants equitable access to the documents unless exclusive arrangements were exceptionally justified.¹⁰⁹

The 2013 iteration modified the PSI Directive.¹¹⁰ The 2013 PSI Directive introduced a general right for third parties to reuse documents that public sector entities make public.¹¹¹ It also modified the provisions on cost recovery to limit charges to the “marginal costs incurred for [the] reproduction, provision and dissemination [of such documents]” in most instances.¹¹² Further, it extended the application of the Directive to categories of organisations previously excluded from its ambit, such as “libraries, museums, and archives”,¹¹³ but it preserved the right of those organisations to a “reasonable return on investment”,¹¹⁴ in establishing access

¹⁰⁴ Open Data Directive.

¹⁰⁵ Open Data Directive, Art. 1(2)(d).

¹⁰⁶ PSI Directive, Art. 3.

¹⁰⁷ PSI Directive, Art. 6.

¹⁰⁸ PSI Directive, Art. 8.

¹⁰⁹ PSI Directive, Art. 11.

¹¹⁰ PSI Directive 2013, Recitals 7 and 8.

¹¹¹ PSI Directive 2013, Art. 3.

¹¹² PSI Directive 2013, Art. 6.

¹¹³ PSI Directive 2013. *See also* Janssen and Hugelier (2013).

¹¹⁴ PSI Directive 2013, Art. 6(4).

costs. Public bodies required to recover costs to fund their operation were also exempted from the requirement to limit their charges to the marginal costs of reproduction and related activities.¹¹⁵

In 2019, the Open Data Directive replaced the PSI Directive, reprising most of its provisions whilst integrating select changes and novelties that build upon and extend the guarantees of openness integrated to the 2003 and 2013 versions of the PSI Directive. One of the most significant changes, for our purposes, is to subject public-funded research data to the full range of reuse requirements that the Open Data Directive, so long as “researchers, research performing organizations, or research funding organizations”¹¹⁶ have made those datasets public through research repositories or other channels.¹¹⁷ The other significant change is to adopt both articles¹¹⁸ and Recitals¹¹⁹ confirming that public sector bodies cannot exercise the *sui generis* database right in a manner that would frustrate the rights to access provided for in the Open Data Directive.¹²⁰

The straightforward conclusion that might be drawn here is that the Open Data Directive has mooted the Database Directive relative to research data and related research outputs. However, the situation is more complex than it might first appear. First, the Open Data Directive is not applicable to numerous categories of datasets that are carved-out from its application.¹²¹ These include those that are subject to IPRs belonging to third parties. If both third-party IPRs and public sector *sui generis* database rights find application to the same dataset, this could re-actualise the right of public sector actors or public funded research to use the *sui generis* database right to preclude downstream use of affected datasets or documents.¹²² This exception could be leveraged in a strategic manner to defeat the intent of the Open Data Directive. One other crucial carve-out applies to identifiable personal data, as defined in EU data protection legislation – insofar as the relevant Union and Member State access-to-information regimes create limitations to the access to such information. Therefore, where research data derived from human participants research is considered to be regulated personal data to which relevant national or Union access-to-information regimes preclude or caveat reuse, the Open Data Directive will not operate to limit the application of the *sui generis* database right.¹²³ Last, it bears mentioning that the Open Data Directive is not applicable to documents that are made public and available for reuse at the discretion of public

¹¹⁵ PSI Directive 2013, Art. 6(2)(a), 6(2)(b).

¹¹⁶ Open Data Directive, Art. 10(2).

¹¹⁷ Open Data Directive, Art. 10.

¹¹⁸ Open Data Directive, Art. 1(6).

¹¹⁹ Open Data Directive, Recitals 54, Recital 61.

¹²⁰ Open Data Directive.

¹²¹ It bears mentioning, however, that numerous categories of data to which “carve-outs” from the application of the Open Data Directive are subject instead to Chapter II of the Data Governance Act. Chapter II of the Data Governance Act applies where public sector bodies make available for reuse certain categories of rights or use restrictions encumber. It establishes standard rules applicable to such reuse, in a similar vein as did the original PSI Directive.

¹²² Open Data Directive, Art. 1(2)(c).

¹²³ Open Data Directive, Art. 1(2)(h).

bodies in a manner that falls “outside the scope of the public task”,¹²⁴ or that fall outside of the relevant national access-to-information regimes.¹²⁵

Second, there are significant definitional challenges in determining the interplay between the obligation to enable reuse on the one hand, and the residual right to preclude extraction or reutilisation on the other. That is: the degree to which public sector entities subject to the obligations in the Open Data Directive can exercise the *sui generis* database right, whilst still fulfilling their obligation to enable the reuse of data (i.e. to refrain from its exercise), remains ambiguous.

Nonetheless, the general trend of the iterative changes made to the PSI Directive is to increase its application to research-related documents and public-funded research data so as to enable their downstream use in a manner that reads down or overrides reliance on private rights therein. There is good reason to imagine that a policy-minded court would interpret the right to reuse in the Open Data Directive as curtailing reliance on the *sui generis* database right. Indeed, the Open Data Directive is but one example of an emergent trend on the part of EU legislators to enact laws privileging both open science and a broad conception of the scientific and public information commons.¹²⁶ Further examination of the contents of the Open Data Directive illustrates this well.

Documents that fall within the scope of the Open Data Directive must be made available in a form that is “open, machine-readable, accessible, findable and reusable.”¹²⁷ Public bodies must respond to requests for access to documents.¹²⁸ Costs charged for access to most documents must be limited to the marginal costs of providing access, and of select other related acts (e.g. anonymising them).¹²⁹ No costs can be charged to access research data.¹³⁰ Public bodies must provide information about the charges applicable to accessing documents¹³¹ and also inform

¹²⁴ Open Data Directive, Art. 1(2)(a). *See also* Art. 1(2)(f).

¹²⁵ This latter limitation is relevant where research-relevant data is made available through national access-to-information regimes (e.g. the public sharing of public health data, population health data, or clinical trial data), but is presumably not relevant where public-funded research data is made available by actors named in Art. 10 according to the access regime directed to research data.

¹²⁶ *See e.g.* van Eechoud (2022).

¹²⁷ Open Data Directive, Art. 5(1). Additional obligations are also applicable to “dynamic data”, which requires these data to be made available “for immediate re-use, immediately after collection, via suitable APIs and, where relevant, as a bulk download.” (*See* Art. 5(5)). These are: “documents in a digital form, subject to frequent or real-time updates, in particular because of their volatility or rapid obsolescence; data generated by sensors are typically considered to be dynamic data.” (*See* Art. 2(8)).

¹²⁸ Open Data Directive, Art. 4. Certain conditions are applicable to this requirement, and certain categories of actors are exempted, such as “public undertakings ... research performing organisations and research funding organisations.”

¹²⁹ Open Data Directive, Art. 6. As discussed in the foregoing sections, some exceptions are applicable to this general rule. For example, libraries are entitled to recover enumerated additional costs, such as those related to “rights clearance” and those related to “data storage.” Public undertakings and other public sector bodies that are required to “generate revenue to cover a substantial part of their costs relating to the performance of their public tasks.” (Open Data Directive, Art. 6(2)(a), 6(2)(c), 6(4)).

¹³⁰ Open Data Directive, Art. 6(6)(b).

¹³¹ Open Data Directive, Art. 7(1), 7(2).

third parties of the redress mechanisms that exist concerning applications for access to documents.¹³²

The Open Data Directive generally prohibits public sector bodies from subjecting third parties' access to documents to restrictive conditions,¹³³ or granting inequitable rights of use to different users.¹³⁴ These limits include express preclusions on granting exclusive rights to documents to select parties¹³⁵ and on imposing discriminatory conditions of access to documents, relative to similar use-cases.¹³⁶

3.5 The Data Act

The European Commission's proposed Data Act, as now drafted, would further read down the scope of application of the *sui generis* database right. Article 35 establishes that the database right would not find application to "databases containing data obtained from or generated by the use of a product or a related service."¹³⁷ If implemented, this should be construed as the near elimination of the database right. The modification is described in the Recitals and explanatory memorandum of the Data Act as directed to ensuring that IPRs do not find application to data that derive from "internet-of-things" (IoT) devices.¹³⁸ Taken alone, this might militate for a narrow interpretation of its breadth of application. In the Impact Assessment Report that accompanies the Data Act, the article is construed first as a clarification detailing that the *sui generis* database right does not find application to data arising from the use of connected devices – as might already be the case without revision because such data arise from pre-existing activities without independent investment – in other words, IoT databases are spin-off databases.¹³⁹ This amendment is also construed as the limited paring down of the database right to preclude such right from generating heightened transaction costs and occasioning reduced potential for competition, and analogised to wider EU

¹³² Open Data Directive, Art. 7(3).

¹³³ Open Data Directive, Arts. 8(1), 8(2). Imposing conditions is presumed impermissible, unless these "are objective, proportionate, non-discriminatory and justified on grounds of a public interest objective" (Art. 8(1)). There is a further, express preclusion on conditions that "restrict possibilities for re-use" or "restrict competition." (Art. 8(2)).

¹³⁴ Open Data Directive, Arts. 8, 11, 12.

¹³⁵ Open Data Directive, Art. 12. Exclusive rights can nonetheless be justified in some circumstances, but these are presumed unlawful, and such exclusive arrangements must be subject to review every three years (Art. 12(2)). These preclusions arise whether the conditions are explicitly discriminatory, or whether the conditions are such as to constructively cause discrimination to arise in practice.

¹³⁶ Open Data Directive, Art. 11. This article contains an express preclusion on imposing discriminatory conditions of use in a cross-border context.

¹³⁷ Data Act (Proposal), Art. 35.

¹³⁸ Data Act (Proposal), Recital 84.

¹³⁹ European Commission. Impact Assessment Report Accompanying the Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act) (2022), at Annex 6.

policy efforts to redraft the *sui generis* database right.¹⁴⁰ Both due to its broad language and its association to the broader effort to limit the negative effects of the *sui generis* database right, the proposed amendment would create wide latitude for courts to further narrow the databases captured by the database right, through a broad construction of the exclusion at Art. 35 of the proposed Data Act.¹⁴¹

4 Policy Considerations for Public Research Databases

Having described general legal and policy considerations relevant to the *sui generis* database right, the Open Data Directive, and select other EU laws, we now consider their application to public biological databases.

Multiple distinct categories of stakeholder engage in the production of public data repositories, including public biological databases. These include research databases funded through public monies, or through a combination of private and public funds, as well as co-created databases that are produced through open science and citizen science initiatives. Some of these repositories are large databases produced to house data from one study or from a network of related studies (e.g. the International Cancer Genome Consortium, UK Biobank).¹⁴² Others are larger-scale repositories that are created to store and enable access to data from multiple unrelated studies, acting as infrastructure at the disposal of researchers that independently choose to deposit data (e.g. dbGaP, the European Genome Archive).¹⁴³ Last, some tools act as “discovery tools” that enable searches to be performed across multiple independent databases or repositories without centralising the queried datasets or dataset metadata (e.g. a Beacon network, CanDIG).¹⁴⁴

The creation of public biological databases often relies on academic research or infrastructure funds from national or international funding agencies, such as the National Science Foundation (NSF), the National Institutes of Health (NIH), the European Union’s Horizon 2020 initiative, and Canada’s Tri-Councils. Other government agencies may also support such activities, including the Defense Advanced Research Projects Agency (DARPA) in the United States. Private charities, the private sector, or public-private partnerships may also support research data generation, curation, and related initiatives. The funding often flows to one or multiple research institutions, government agencies, or consortia, which support, through the provision of data, intra- or extra-mural biomedical research. These

¹⁴⁰ European Commission. Impact Assessment Report Accompanying the Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act) (2022).

¹⁴¹ See also Keller. A vanishing right? The *Sui Generis* Database Right and the proposed Data Act. Kluwer Copyright Blog. Available from: <https://copyrightblog.kluweriplaw.com/2022/03/04/a-vanishing-right-the-sui-generis-database-right-and-the-proposed-data-act/>.

¹⁴² The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium (2020); Sudlow et al. (2015); Zhang et al. (2011).

¹⁴³ Tryka et al. (2014); Lappalainen et al. (2015); Freeberg et al. (2022).

¹⁴⁴ Dursi et al. (2021); Rambla et al. (2022).

categories of repositories differ in their organisational structure, their purpose, and in the nature of the information stored.

The benefit of public data lies in its potential for reuse. Therefore, the default operation of law – through the *sui generis* database right – granting exclusive IPRs on public datasets conflicts with the rationale that justifies their creation. This could lead to the automatic creation of private rights in public datasets, that run contrary to the intentions of public research institutions, public funding agencies, and self-organised open science communities.

In the following section, we apply the tests discussed above to determine when the *sui generis* database right might apply to public biological databases. The subsequent section determines the conditions according to which the creation of platforms used to discover data or to submit queries to existing public biological databases will breach the *sui generis* database right.

4.1 Does the *Sui Generis* Database Right Apply?

4.1.1 *Barriers to the Recognition of the Sui Generis Database Right that Relate to Public Biological Databases Meeting Its Preconditions*

Three reasons support the conclusion that the database right is not applicable to public biological databases, which relate to the investment test and the conditions according to which such databases are compiled. First, the database right will not apply to databases that generate their own data as per *British Horseracing Board* and related decisions.¹⁴⁵ Because the production of such databases is not subject to an investment independent from that made to create the concerned data, it is not subject to the *sui generis* database right. That is, funds (often public) are invested to perform the first-instance generation of research data, rather than to curate pre-existing data into databases.¹⁴⁶ While the CJEU has not directly addressed the application of the *sui generis* database right to public research databases, the national courts of Member States have. National courts have allowed the private sector participants in public-private partnerships to benefit from the *sui generis* database right, whilst denying the right to the public sector partners. Thus, national courts have sometimes found that institutions/agencies that obtain public funding to create a database, or to generate raw data, do not benefit from the *sui generis* database right.¹⁴⁷ Second, research databases are often created through the principal research activities of the consortium or its constituent research organisations, and therefore no independent investment is made in the production of such databases

¹⁴⁵ *British Horseracing Board; Latvia*.

¹⁴⁶ Beunen (2007), pp. 121–122, 126; Compagnucci (2020). Certain authors, including Beunen, contend that it is unclear whether databases that result from scientific research activities are constituted from data that is created or data that is collected. This determination is material because created data is presumed not to benefit from *sui generis* data protection, whilst collected data can benefit from the *sui generis* database right if the other preconditions to the recognition of the right are satisfied.

¹⁴⁷ Derclaye and Husovec (2021).

(i.e. these are spin-off databases).¹⁴⁸ Third, the independent investment made in the creation of a public biological database likely carries *no investment risk*, which is a precondition to the database right being recognised. That is, the data were produced through funding directed to assembling databases and other research outputs that act as pure public goods – the output data were not intended to be translated into profits, and therefore no associated investment risk exists.

One notable case that might counterbalance the foregoing analysis is the CJEU holding in *Directmedia*. In *Directmedia*, a university invested EUR 34,900 in compiling a list of the most common German poems featured in existing anthologies, which required archival research and statistical analysis.¹⁴⁹ This investment was considered to be subject to the protection of the *sui generis* database right. However, it is worth considering that in this instance, significant resources were exceptionally dedicated to the compilation of data from existing sources (rather than to data generation) – and the question of whether or not the concerned database was subject to protection was not itself referred to the court.¹⁵⁰ The court was asked to produce a holding on the issue of infringement, alone. This holding could temper our conclusion that public-funded and research-funded efforts to compile databases absent a profit motive are less subject to the *sui generis* database right than private-funded efforts that envision profit, or public-funded efforts that envision cost-recovery (e.g. through the levy of access fees).

Certain ambiguities therefore remain. It is unclear whether public-sector research funding arrangements that enable funded parties to benefit from cost-recovery mechanisms qualify for *sui generis* database protection. Second, the threshold of additional investment required to translate a created, spin-off, public-funded, or otherwise presumptively unprotected database into a protected database is difficult to ascertain. It is also unclear how a recipient of research funding could invest additional funds and effort at its own risk to translate a non-protected deliverable into a protected object that is the result of additional, risk-bearing investment.

¹⁴⁸ Prior to the *British Horseracing Board* tetralogy of cases, support for the “spin off theory” (stipulating that “spin-off databases” which are produced serendipitously through the main activities of the database creators absent additional effort or investment should not qualify for the protection of the *sui generis* database right) was not unanimous in national jurisprudence. For example, courts in the Netherlands were divided on this issue, whilst courts in France often reject “spin off theory” and accorded rights in databases that created through the main activities of the database creator absent additional investment. See Hugenholtz (2001); Derclaye (2004); Derclaye (2012). In this latter article, Derclaye notes that national courts in EU Member States have taken quite some time to internalise and apply the CJEU holding in the *British Horseracing Board* tetralogy, but that judgments which are consistent with the CJEU holdings in the *British Horseracing Board* tetralogy are being issued with increasing frequency. In other words, national courts are increasingly becoming apprised of the CJEU’s denial of the database right to spin-off databases, and issue judgments consistent with this outcome (that is, with the CJEU’s tacit confirmation of the spin-off theory). However, perfect implementation of the CJEU holdings that implicitly confirm the spin-off theory remain lacking because numerous national court judges still remain unaware of them.

¹⁴⁹ *Directmedia*, at paras. 9–12.

¹⁵⁰ *Directmedia*, at paras. 23–24.

4.1.2 *Barriers to the Recognition of the Sui Generis Database Right Arising from the Definition of a Database “Maker”*

The identification of an appropriate database “maker” that benefits from the *sui generis* database right can be unclear in research collaborations. For example, the right vests in “the person who takes the initiative and the risk of investing”,¹⁵¹ leaving ambiguous whether the right could vest in actors that initiate the database or those that bear the risk inherent in creating the database, alone.¹⁵² Alternatively, it is possible that both conditions must be met,¹⁵³ but it is unclear whether each collaborator must meet *both* requirements, or whether it is sufficient for separate collaborators to meet each of the conditions (so as to hold either separate or joint rights).¹⁵⁴ Recent case law does appear to suggest that one actor must perform the initiation of database creation, and also bear the associated investment risk, to benefit from the database right.¹⁵⁵

Equally ambiguous are the categories of risk sufficient to satisfy the “investment risk” criterion.¹⁵⁶ Some legal scholars conclude that the Database Directive refers to the “organisational” risk engaged in creating the database, rather than the financial or reputational risk inherent in its creation.¹⁵⁷ This interpretation implies that “pure contractors” (i.e. persons that hire third parties to produce a database) could not benefit from the *sui generis* database right.¹⁵⁸ We disagree, because such limitation of the categories of “risk” to “organisational risk” does not flow from the text of the Database Directive. Instead both the historical origin of the right¹⁵⁹ and the

¹⁵¹ Beunen (2007), pp. 154–156.

¹⁵² Beunen (2007), pp. 148–151. See also Schellekens (2011), p. 624. Though Recital 41 posits both conditions to be cumulative requirements, Beunen notes that numerous authors, and select Member State implementations of the Database Directive, have considered the “initiative” criterion to be non-binding and have not incorporated such precondition to applicable domestic legislation.

¹⁵³ Beunen (2007), pp. 146–152; Schellekens (2011). Schellekens favors the cumulative interpretation (p. 625).

¹⁵⁴ The court in *Latvia* and that in *Innoweb* appear to frame these requirements as being cumulative. See *Latvia* at para. 22 and *Innoweb* at para. 36.

¹⁵⁵ The court in *Latvia* and that in *Innoweb* appear to frame these requirements as being cumulative. See *Latvia* at para. 22 and *Innoweb* at para. 36. See also *ABRvS 29 April 2009, n 07/786, AMI 2009-6* (College B&W Amsterdam/Landmark; m. nt. M. Van Eechoud). This case is discussed at length in: Derclaye and Husovec (2021), p. 8 and in Buri (2012), pp. 24–27.

¹⁵⁶ Beunen (2007), pp. 151–153. Beunen borrows the tripartite taxonomy that Seignette proposed prior to describe the potential categories of risk that could lead to the recognition of a *sui generis* database right.

¹⁵⁷ Beunen (2007), pp. 151–153, 157.

¹⁵⁸ Beunen (2007), pp. 148–150, 152. According to Beunen, this is the case because the “risk” that must be made to benefit from the *sui generis* database right is not the financial investment in the “production” of a database, but rather the true organisational risk entailed in the creation of the database.

¹⁵⁹ That is, its attempt to preserve the English “sweat of the brow” standard for copyright protection despite the elimination from EU copyright law.

exclusion of subcontractors¹⁶⁰ at Recital 41¹⁶¹ favor an interpretation that accepts pure or principally financial risk as sufficient to give rise to the database right.¹⁶² Recital 40 appears to contemplate the expenditure of “financial resources” and the “expending of time, effort and energy” as discrete and valid forms of risk-taking that could give rise to the right independent from one another.¹⁶³ Courts appear to consider that one organisation must take both the “initiative” and the “risk” of investing, in cumulation, to benefit from the *sui generis* database right. Conversely, it is not required for one singular organisation to bear a specific category of risk to benefit from the *sui generis* database right (i.e. to expend resources that are explicitly of a financial, organisational, or other category in the production of the database).¹⁶⁴

Based on this discussion, the potential for individual or institutional members of research consortia to benefit from the database right may be restricted in practice, depending on the interpretation of “database maker”. Research consortia will oftentimes not meet the definition if all potential rights-holders must meet *both* the precondition of initiation and risk, or if the risk must be organisational in character. Conversely, if a more liberal interpretation is applied, collaborators within research consortia, either individually or as a group, may benefit from the database right.

Interpretive debates notwithstanding, the policy objective of the Database Directive is to create incentives to invest in database compilation. This suggests that courts will favor a broad definition of database maker, including network-based, market-based, or firm-based production, rather than recognising elements contingent on specific organisational forms or financial arrangements. In practice, the “investment” test will be more determinative of whether a *sui generis* database right is recognised, and the apportionment of roles and responsibilities between research collaborators will be less so. This interpretation accords with the investment test and the public policy considerations that underpin the *sui generis* database right; the definition of database maker is tangential.

4.1.3 *Barriers to the Recognition of the Sui Generis Database Right Arising from the Structure of Research Collaborations and Research Consortia*

It is often unclear how national law apportions joint ownership of property rights, including IPRs. Thus, singular institutions within research consortia, or groups of unincorporated researchers, could potentially hold joint *sui generis* database rights

¹⁶⁰ Beunen (2007), pp. 148–151. According to Beunen, this is the case because Recital 41 excludes subcontractors from being considered the “maker” of a database. However, Beunen (p. 151) does consider that this is might not be a *strict* preclusion, but rather that a subcontractor will not often, on the facts, be considered the principal actor that has made the “substantial investment” in the creation of a protected database.

¹⁶¹ Database Directive, Recital 41.

¹⁶² Schellekens (2011), pp. 623, 625. Schellekens places reliance on the language of the CJEU in *Oy Veikkaus Ab* to conclude that pure financial risk is one of the categories of risk that creates an entitlement to the *sui generis* database right.

¹⁶³ Database Directive, Recital 40. The court in *Latvia* frames their analysis in a similar manner.

¹⁶⁴ *Latvia* at para. 22; *Innoweb* at para. 36.

in the public research databases created via joint efforts. Whether these rights can be exercised by a singular consortium member, or must be exercised in a collective manner, might depend on the contents of applicable national law. The potential for conflicts in interpretation between the national property laws of each EU Member State give rise to further uncertainty.

Application of the database right also leaves indeterminate the rights of non-EU participants in collaborations, as the Directive restricts the *sui generis* database right to “nationals of a Member State or who have their habitual residence in the territory of the community”¹⁶⁵ and “companies and firms formed in accordance with the law of a Member State and having their registered office, central administration or principal place of business within the Community.”¹⁶⁶ It is unclear how EU courts would consider the entitlement of non-EU collaborators to the joint *sui generis* database right arising from an international research collaboration.¹⁶⁷ In transcontinental collaborations, it is conceivable that EU beneficiaries could license the *sui generis* database right to their non-EU collaborators, frustrating the intent of the Database Directive to restrict benefits to EU nationals or residents.

Moreover, it remains unclear whether a research collaborator who benefits from the *sui generis* database right could license such rights to collaborators who do not, as part of a consortium agreement or other contractual agreement. Contrary to the Database Directive, such licensing would enable collaborators that produce databases as part of their existing activities to benefit from the *sui generis* database right.

A further complication is the application of the Database Directive to collaborations in which the majority of the investment and attendant risk is borne by parties that do not qualify for the database right. For example, if an unincorporated public-private partnership or an international effort invests in the creation of a database, and the majority of the investment risk is borne by the non-EU collaborators, it is unclear whether such a collaboration would entitle the qualifying participants to benefit from the *sui generis* database right. Indeed, some legal scholars contend that the text of the Database Directive, independent of national laws, is ambiguous about the potential for joint *sui generis* database rights to arise in a database, the creation of which results from a collaborative effort. Other authors have attempted to resolve the question about the potential for joint ownership of *sui generis* database rights in considering the legal nature of the *sui generis* database right itself (i.e. the text of the Database Directive), rather than the applicable Member State property law rules governing how property rights arise and

¹⁶⁵ Database Directive, Art. 11(1).

¹⁶⁶ Database Directive, Art. 11(2).

¹⁶⁷ Lipton (2009), pp. 654–655. Lipton considers the circumstances in which a collaborative effort to create a database that could benefit from *sui generis* database protection involves both EU persons or entities that could benefit from the *sui generis* database right and non-EU persons or entities that are not entitled to benefit from the *sui generis* database right. It is not clear if this would lead to a collective *sui generis* database right to arise that also benefits non-EU collaborators, whether a right benefiting exclusively the EU collaborators would arise, or whether no *sui generis* right would arise at all. According to Lipton, it is also unclear how material the proportion of EU collaborators to non-EU collaborators in creating a database is to this determination.

are apportioned.¹⁶⁸ Lipton contends that the language of the Database Directive creates ambiguities for the potential for joint *sui generis* database rights to arise in a database, the creation of which results from a collaborative effort.¹⁶⁹ Some legal scholars¹⁷⁰ consider this question relative to the non-binding definition of “database maker” that is provided at Recital 41 of the Database Directive, and the inconsistent Member State implementations.¹⁷¹ This line of inquiry attempts to determine the minimum preconditions that one collaborator in a collective endeavor must satisfy to meet the definition of “database maker”, as discussed above.

4.2 Can the Database Right Limit Use or Reuse of Public Research Data?

In the previous section, we demonstrated that public biological databases will seldom be subject to the *sui generis* database due to the application of the investment test and the “created/collected” distinction, but difficulties might also arise from the definition of the database maker and the joint allocation of rights between collaborators.

In circumstances where a *sui generis* database right is recognised in a database, we now consider the categories of downstream data use that the database right might preclude. We focus on the creation of data discovery platforms, search engines, and other, similar tools that enable pre-existing external databases to be queried and accessed with greater facility.

¹⁶⁸ Lipton (2009), pp. 654–655.

¹⁶⁹ Lipton (2009), pp. 641, 646. Lipton notes that Art. 11(1) of the Database Directive potentially anticipates the potential for multiple individuals to hold the *sui generis* database right in a joint capacity. Lipton draws this inference from the use of the plural in Art. 11 to describe the preconditions for a database to benefit from the *sui generis* right: it “... shall apply to database whose makers or rightholders are nationals of a Member State” Indeed, the Directive’s use of the singular to refer to both the protected database, and the concerned right, whilst using the plural to refer to the prospective rightholders, supports Lipton’s contention that a singular *sui generis* right applicable to a single database could arise in a joint capacity.

¹⁷⁰ Beunen (2007), p. 154. Beunen notes that national Member State jurisprudence has recognised the potential for joint ownership of the *sui generis* database right, shared between non-incorporated groups such as associations. This example is a close analogue to the biomedical research consortium, and suggests that it is perhaps appropriate for a consortium, despite the lack of legal personhood, to be considered as a singular “database maker” and invested with a singular, joint *sui generis* database right.

¹⁷¹ Beunen (2007), pp. 154–156. Beunen describes the academic literature on “joint” ownership of the *sui generis* database right and enumerates a number of live controversies amongst the relevant authors. Such controversies include, first, whether the text of the Database Directive and its respective national implementations permit joint ownership of the *sui generis* database right to arise through the operation of law (*i.e.* as opposed to joint ownership arising through the contractual assignment of the database right). Second, authors differ as to which criteria of the definition of “database maker” that is provided in Recital 41 of the Database Directive *each* contributor to a database must satisfy to be entitled to the collective *sui generis* database right. These difficulties arise because the nature of the “database maker” is not defined in the binding text of the Database Directive, but rather in the non-binding Recital 41. Therefore, different Member States have settled on divergent implementations of the preconditions to be considered a “database maker” (Beunen, p. 155). Moreover, it appears unsettled law whether a collective *sui generis* database right would be accorded to each of the contributors to a database if each satisfied one, but not all, of the preconditions to the right (*i.e.* initiative and risk) – compared to the circumstance in which each contributor satisfies each of the preconditions (*i.e.* initiative and risk) (Beunen, p. 155).

Do the creation and operation of platforms that facilitate data discovery or data access have the potential to infringe the Database Directive? Platforms host information about available datasets, known as metadata, and triage the metadata to help researchers find information that is relevant to their intended research activities, hosted on other pre-existing repositories. For example, a platform could enable multiple repositories of biological data to be queried for features that are of research interest. Queries could be based on genetic variant, phenotype, pathology, or other elements of clinical metadata, and formatted using standardised ontologies.¹⁷² These platforms could act as pure “discovery” tools, enabling researchers to search for relevant dataset-level information from across multiple restricted-access (controlled) repositories. Users then proceed to negotiate access to the datasets of interest with their individual data custodians. Alternatively, platforms enabling data “access” could centralise the direct download links to multiple external datasets that are of common research interest to downstream data users. Still other technologies could bring together some features of “discovery” and “access” tools, for example, those that produce aggregate data from dataset-level metadata or from record-level data within protected datasets. These could add value to the datasets used to produce them, or could render them obsolete.

4.2.1 *Creating Downstream Databases from Existing Databases and the Sui Generis Database Right*

The direct reutilisation of a pre-existing database as part of a second, downstream database could infringe the *sui generis* database if elements of one public biological database are integrated to a downstream database that fulfils a similar purpose. In *Directmedia*, the CJEU held that broad categories of actions could be held to constitute “extraction” or “reutilisation”, infringing the *sui generis* database right. These acts are not limited to direct “copying” of databases through technological duplication but also to their labor-intensive manual reconstitution in full or in part.¹⁷³ Therefore, in producing downstream databases that derive considerable content from other existing databases, the operative analysis is not directed to the categories of “actions” performed to create copies of the affected databases. Rather, it is directed to whether their outcome constitutes the “extraction” or “reutilisation” of a substantial part of a protected database. This requires the application of the test elaborated in *Latvia*, which seeks to determine, first, whether the extraction or reutilisation of a substantial part of a protected database has occurred, and second, whether this extraction or reutilisation creates a “significant detriment, evaluated qualitatively or quantitatively, to the investment.”¹⁷⁴ In performing this evaluation,

¹⁷² Ontologies are technical vocabularies that define a set of related concepts within a discipline or topic area and clarify the nature of and the relations between each such concepts. These are often used to standardise and formalise discussions within a specialised discipline and eliminate vagueness or ambiguities arising from the informal expression or comparison.

¹⁷³ *Directmedia* at paras. 22–60.

¹⁷⁴ *Latvia* at para. 39.

the CJEU in *Latvia* also considers the balance between precluding the “potential risk to the substantial investment of the maker”¹⁷⁵ and privileging the entitlement of third parties to access information and/or to translate it into downstream innovations.

The creation of a derived database that replicates a pre-existing database – or brings content from multiple such databases together – therefore does not necessarily infringe the *sui generis* database right. First, we determine whether the derived database extracts and reutilises a substantial portion of the original database. Second, we determine whether or not this detracts a great deal from the potential for the rights-holder to benefit from their investment. Third, where such a risk arises, we evaluate its severity relative to the harm to functioning markets in information that would be occasioned in upholding the right. This means evaluating the value of repurposing such information, or of a value-added service that relies on it.

A derived database that acts as a strict replica of a pre-existing, protected database will likely violate the *sui generis* database right. Conversely, a derived database that brings together information from one or multiple databases without detracting from the value of the original database (for example, because its function differs, or it is directed to a different audience), would not breach the database right. Databases that extract and/or reutilise a “substantial part” of one or multiple databases in a manner that deprecates the value of the original databases will not necessarily breach the *sui generis* database right, so long as these produce value that is independent from that of the original database, sufficient to justify the risk to the investment.¹⁷⁶ Nonetheless, the presence or absence of a risk to a protected investment remains the main criterion that is used to determine whether or not the *sui generis* database right is infringed.¹⁷⁷

4.2.2 *Creating Platforms and Tools Enabling Data Discovery, Analysis, and Access that Aggregate Data from Multiple Databases or Repositories*

A similar analysis determines whether or not the integration of data, metadata, download links, and the like from other public biological databases into downstream discovery tools or search engines breaches the *sui generis* database right.

The CJEU has explicitly addressed the relationship between value-added search platforms and the database right. In *Innoweb B.V. v. Wegener I.C.T. Media B.V. (Innoweb)*, the CJEU considered whether a “meta search engine”¹⁷⁸ breached the database right. The court held in the affirmative because the meta search engine reutilised a substantial portion of the underlying databases searched.¹⁷⁹ The meta

¹⁷⁵ *Latvia* at para. 44, *see also* paras. 46–47.

¹⁷⁶ *Latvia* at paras. 44–46.

¹⁷⁷ *Latvia* at paras. 44–46.

¹⁷⁸ A meta search engine is a tool that aggregates the results of searches that other search engines performed. Case C-202/12 *Innoweb B.V. v. Wegener I.C.T. Media B.V. and Wegener Mediaventions BV* ECLI:EU:C:2013:850 at para. 8 (referred to as *Innoweb*).

¹⁷⁹ *Innoweb* at paras. 48, 54.

search engine, GasPedaal, accepts queries for car advertisements according to select filters and search terms. Once a search is made using the platform, the meta search engine runs the search directly via underlying search platforms. GasPedaal then presents the search results to the end user, integrating information in a similar order to the underlying search platforms queried. End users are therefore not required to consult the underlying platforms.¹⁸⁰ The CJEU stated: “to ensure that the person who has taken the initiative and assumed the risk of making a substantial investment in terms of human, technical, and/or financial resources in the setting up and operation of a database receives a return on his investment.”¹⁸¹ GasPedaal infringed the database right, because it rendered irrelevant the search engines from which the presented information was drawn; it deprived such platforms of income.¹⁸² Further, the meta search engine, despite providing additional value by aggregating the results of multiple search engines, operated in a similar fashion to the source search engines. It presented search queries in a comparable manner, and performed searches in real time.¹⁸³ In sum, the court concluded that these features rendered the GasPedaal meta search engine “close to the manufacture of a parasitical competing product” and therefore to infringe the database right.¹⁸⁴

The CJEU considered meta search engines again, on different facts, in *Latvia*. The search platform concerned indexed the “meta tags” of numerous websites hosting job advertisements, that such websites provided to make their content discoverable to general search engines. In contrast to *Innoweb*, the search engine in *Latvia* did not reuse the native search engines of the websites from which it drew its content.¹⁸⁵ Instead, it first indexed the content of the websites that hosted the job advertisements and then provided deep links to those websites’ primary copies of the job advertisements.¹⁸⁶ The impugned search engine added functionality in bringing together in one place the results of queries that would otherwise need to be performed on numerous independent search engines. Unlike GasPedaal, it did not constitute a parasitic imitation of the search engines used to produce it. GasPedaal did little but run searches through the existing search engines, simultaneously depriving the underlying sites of usership and revenue.¹⁸⁷

¹⁸⁰ *Innoweb* at paras. 9–11, 25, 54.

¹⁸¹ *Innoweb* at para. 36. The court further considers that the holder of the *sui generis* database right is protected from acts that have the effect of “depriving him of revenue which should have enabled him to redeem the cost of his investment.

¹⁸² *Innoweb* at paras. 40, 42.

¹⁸³ *Innoweb* at para. 54.

¹⁸⁴ *Innoweb* at para. 48. See also European Commission (2018). The authors of this study describe that part of the original efforts to implement the Database Directive intended the EU-wide harmonisation of the tort of slavish imitation, or tort of parasitism. The *sui generis* database right was adopted in part as the effort at harmonisation proved impossible. It is therefore sensible that courts should consider such policy imperatives in determining whether or not the *sui generis* database right has been breached.

¹⁸⁵ *Latvia* at para. 19.

¹⁸⁶ *Latvia* at paras. 10–11.

¹⁸⁷ *Innoweb* at para. 48.

4.2.3 Conclusion as to the Effects of the Database Right

Our general position on the application of the *sui generis* database right to public biological databases is the following. There is strong, but not unqualified, support, for the position that the *sui generis* database right would not apply to most public-funded consortium-produced public biological databases. The strongest argument in support of this claim is that such efforts often “create” rather than “collect” data, which does not create an entitlement to the *sui generis* database right. Further, recent case law provides heightened support for the position that the *sui generis* database right does not find application to public-funded databases due to a lack of associated investment risk, though this position remains contentious. If the database is produced as the incidental result of pre-existing research or clinical activities, it is almost certain that no *sui generis* database right will be recognised. Last, different actors taking the initiative and bearing the risk of investing in the creation of a database could lead to their disqualification from database protection due to the cumulative nature of such criteria. Our general conclusion is that researchers operating as a consortium to produce a database will often be deprived from the protection of the *sui generis* database right, though the preclusion is not strict and absolute.

In those residual circumstances in which the *sui generis* database right does find application to a public biological database, the prospect for downstream use of that database to infringe the database right is restricted. First, some authors contend that because the framing of infringement in *Latvia* requires that the downstream use create a risk to the fruition of the protected investment, public-funded research deliverables would struggle to overcome the infringement test even where subject to the database right. Second, there is a high likelihood that either derived databases or discovery-and-access tools that add value to existing databases would not infringe the *sui generis* database right even if these created a risk to the realisation of the investment in a protected database. This is the case because of the application of the balancing test introduced in *Latvia*.

The final residual consideration is the application of the Open Data Directive and the Data Governance Act in these circumstances. These also contain provisions that enable reuse of research data and public sector information and limit the potential to invoke the *sui generis* database right. Though relevant, it is not clear how broad the right of reuse anticipated in these acts is, and there is significant potential for research data not to be subject to these laws due to the carve-outs anticipated for certain categories of data. Therefore, these instruments are anticipated to be less impactful than the considerations raised in the preceding paragraphs.

The caveat to our position is the holding in *Ryanair*. This holding enables organisations that release databases that are not subject to the *sui generis* database right to use strong contractual protections or digital rights management (DRM) technologies to condition the use and reuse of those databases. In using such protections, organisations need not respect the minimal rights of reuse that the Database Directive anticipates for those databases that are subject to the *sui generis* database right. Therefore, it is possible that the producers of public biological databases could assert strong contractual protections – or use digital rights

management technologies – to prohibit the reuse of their databases, where the *sui generis* database right does not find application. However, the use of such conditions in the public sector, at least, is often precluded by open access policies of funding agencies and research institutions.

The lack of clarity as to whether the database right applies to research databases and associated information, including metadata, annotations, and ontologies, combined with the potential for unintentional infringement, could deter third parties from producing data discovery platforms, data access platforms, enriched datasets, and other value-added derivatives. Legal uncertainty may negatively impact open science initiatives that rely on data and tools contributed by third parties as well as ongoing analysis and reutilisation of data. Reform is therefore recommended.

4.2.4 Conclusion as to the Benefits and Consequences of the Database Right

The legal uncertainty of the database right, compared to copyright, is exacerbated because of its broad scope and lack of time limitation. A more limited right might still compensate database producers for the high cost of first production that places them at a competitive disadvantage relative to other participants in the same market.¹⁸⁸ It might stimulate investment in the production of intangibles in the form of databases. Without some protection, the market might under-appportion the creation of such goods because they are non-excludable, precluding prospective creators from recognising profits on their investment in creation.¹⁸⁹ However, a more certain and time-limited right would decrease the risk of exclusion inhibiting the creation of downstream value. Under the current regimen, rights-holders may engage in rent-seeking behaviors, commanding high prices to access IPR protected intangibles that could otherwise be shared or reproduced at zero or minimal cost. In the case of the database right, this risk is exacerbated because recent policy analyses conclude that there is limited need to introduce legal rights to incentivise the production of databases – database producers have increasing incentives to do so as other economic activities become increasingly data-reliant.

It is apparent from our discussion that the CJEU and national courts have attempted to optimise the scope of protection, with respect to both scope and subject matter. They have limited the scope of the database right in step with the emergence of a mature information society, in which numerous incentives to generate valuable data and datasets exist. They recognise the economic value of enabling access to data and datasets, limiting exclusive rights that purportedly incentivise their production and dissemination.¹⁹⁰

Our discussion highlights the interpretive ambiguities and potential inflexibilities of IPRs, such as the Database Directive. Such IPRs are problematic because the proliferation of differing statutory interpretations risks real or perceived legal non-compliance. Ambiguities in the breadth of application of IPRs discourages

¹⁸⁸ Hadfield (1998).

¹⁸⁹ Hadfield (1998).

¹⁹⁰ European Commission (2018), pp. 25–32.

investment in the production or disclosure of intangibles for fear that these will not be protected. It also encourages the reticent use of available intangibles to avoid potential infringement.

Unfortunately, opportunities for legal reform are resource-intensive and slow. They require action by the judiciary or legislators. It is therefore difficult to ensure that the scope of IPRs remains context-sensitive and attuned to economic, societal, and technological transformation.

5 Proposed Amendments to the *Sui Generis* Database Right

5.1 Proposals to Reform the *Sui Generis* Database Right

Academic commentators and the formal evaluations of the Database Directive have proposed a number of reforms, up to and including the elimination of the database right. The arguments against repeal of the right are first, that its elimination could expose firms whose database rights have financial value to financial loss. Second, such action could potentially constitute a breach of the right to intellectual property, which is established in EU human rights law.¹⁹¹ Furthermore, such action could cause firms to limit their production of EU-sourced intangibles (or purchase of rights therein) for fear of the subsequent elimination or aggressive modification of the IPRs. Finally, the Database Directive not only created the database right but also articulated the compromise position between the continental and the UK approach to copyright protection. Despite the departure of the UK from the EU, the continued harmonisation of the database right and copyright in both jurisdictions creates economic advantages for firms resident in both jurisdictions.¹⁹²

A second reform proposal is the implementation of a registration requirement and/or a notification requirement as a precondition to asserting the database right.¹⁹³ This would, for instance, require firms that assert their database right to provide such notice on their database(s) and require firms to register the database's digital object identifier (DOI) in a public register. These mechanisms alleviate some legal ambiguity, because they give notice to third parties that extract or reutilise data from a database. Conversely, the absence of rights assertion provides greater certainty regarding freedom to operate. The principal disadvantage of this approach, however, is that it could entrench the economic advantage of large relative to smaller firms, because legally sophisticated firms are more likely to take action on notification and registration.

Other potential reforms include the implementation of novel exemptions to the database right, for example, an exemption for commercial and non-commercial research that extracts and reutilises data and an exemption that enables TDM. The

¹⁹¹ European Commission (2018), p. 126.

¹⁹² European Commission (2018), pp. 131–132. The study, however, notes that disparate national implementations of the *sui generis* database right, and of the exceptions thereto, might frustrate this harmonising effect.

¹⁹³ European Commission (2018), p. 141.

TDM exemption as currently construed enables database right-holders to opt out of its application,¹⁹⁴ creating uncertainty for users, and limits its application to the extraction prong of the database right. Numerous methods of performing TDM are algorithmic, reducing the ability of users to determine if an extracted or reutilised database is subject to an opt-out notice.¹⁹⁵ It also limits its application to extraction and not reutilisation, diminishing its usefulness even where it does find application.¹⁹⁶ Broad and clear exemptions would provide greater legal certainty for users, including users of public research databases.¹⁹⁷

A further issue with the database right that needs to be addressed is the information asymmetry that inhibits rights-holders and downstream users of protected databases from appreciating whether an act constitutes the infringement of the *sui generis* database right with equal precision. The investment test, which is the principal mechanism used to determine if the right arises and, since *Latvia*, to assess its infringement, relies on information about the existence of a protected investment that is not available to the potentially infringing party.¹⁹⁸ A solution may be to mimic copyright exemptions, which include fair use and its subcategory of “transformative fair use” in US copyright law. The test for transformative fair use considers whether creative effort has transformed the object of copyright protection into a novel creative work. Applied to the database right the exemption might consider whether a “substantial investment in terms of human, technical and/or financial resources” creates a new database or other derivative output that is distinct from the protected database, so as to constitute transformative fair use.¹⁹⁹ This is similar to the test introduced in *Latvia*. However, where the *Latvia* test uses information available to the right-holder to determine whether the right is breached,

¹⁹⁴ Albeit not relative to downstream scientific research use by research organisations and cultural organisations – in this case, the right is unqualified. The definition of research organisations suggests that organisations performing principally commercial research are excluded. See Digital Single Market Directive, Arts. 2(1), 3.

¹⁹⁵ Margoni and Kretschmer (2022), p. 687. Margoni and Kretschmer discuss the TDM exception in the context of EU innovation policy and artificial intelligence regulation. Their conclusions dovetail one of the major themes of this article: that the transaction costs and compliance costs of navigating the *sui generis* database right and related legal instruments overshadow the benefits of its well-nuanced substantive regime.

¹⁹⁶ Digital Single Market Directive, Arts. 3, 5.

¹⁹⁷ European Commission (2018).

¹⁹⁸ Derclaye and Husovec (2021), p. 11.

¹⁹⁹ See e.g. Heymann (2008). See also Liu (2019). Heymann illustrates different conceptions of the “transformative” prong in copyright law’s transformative fair use test throughout history. E.g. early jurisprudential “economic” articulations consider the transformative character of the purpose in creating a novel output, such that the test privileges creators that vie for the creation of new works. Conversely, later jurisprudence, and Heymann’s own proposed reformulation of the test, increasingly privilege the external interpreter’s determination of the output work’s independent character. In creating a “transformative” use exception to the *sui generis* database right, it is worthwhile to consider the appropriate interpretive framework to determine whether a “transformation” has occurred or not. In our view, either of the foregoing approaches might be appropriate. The former encourages third parties to invest in the creation of derivative datasets and related outputs in privileging their “transformative” acts; the latter incentivises the creation of derivative datasets and related outputs that create appreciable added-value or are functionally distinct from the originating material, thus balancing the protection of investment against the protection of value-added database derivatives.

the “transformative fair use” test would use information available to the downstream user to assess whether or not infringement had occurred.

A fair use exemption would continue to provide database right-holders with a similar breadth of protection but enable the potentially infringing party to objectively assess whether a derivative output that entails a substantial investment is sufficiently distinct from the initial database to meet the standard of “transformative fair use.” The principal benefit is that the test used to determine if a database is protected and if a prospective derivative infringes the original right is assessed on the basis of information that is available to all parties.

Finally, the Database Directive could be replaced by a regulation that harmonises the database right throughout the EU²⁰⁰. This could perform the explicit clarification of the relationship between the Database Directive that intends to protect investments in database-production, and later EU public policy efforts that are directed to guaranteeing the downstream openness of both public-funded and private-funded data, such as the Open Data Directive, the Data Governance Act, and the Data Act. This effort would be welcome because the policy priorities of EU legislatures have shifted from incentivising the production of databases to incentivising the open release and reuse of data. This shift reflects changes in the market for information and downstream products/services, rather than a shift in the objectives of the European Commission in information policy.

Though these possibilities are appealing, the transaction costs inherent in interpreting the boundaries of the *sui generis* database right and negotiating access rights, and the continued potential for litigation costs or civil penalties to be imposed on good-faith downstream users of data would remain even where the right was amended or its ascription was mediated through a registration requirement or other formality. It is dubious that the database right creates value that equals or surpasses the transaction costs and chilling effect that its ambiguous boundaries impose. For these reasons, and those stated below, we urge the repeal of the *sui generis* database right.

6 Repealing the *Sui Generis* Database Right

Despite potential challenges, we contend that the *sui generis* database right should be repealed. While minor modifications and interpretations have the advantages discussed above, the evaluation and administration of the *sui generis* database right creates considerable costs for database creators, database users, and for courts. Such costs are incurred in developing contracts to manage and to assign the respective database rights of collaborators, and that inhere in both upstream and downstream users of shared data and databases. Costs are also incurred in hiring legal advisors and other relevant experts to determine the probable application of the *sui generis* database right to particular factual circumstances, on the part of both prospective claimants and prospective users of potentially protected databases. Parties must continue to bear the costs of legal ambiguities that lead to disputes or to unintended

²⁰⁰ Derclaye and Husovec (2021), p. 13.

outcomes (e.g. unanticipated infringement of the *sui generis* database right, or reliance on a right that courts deem not to exist).

The costs of the *sui generis* database rights are not limited to those that database producers and data users must bear. The administration of the database right, and the negative externalities that it generates, can create costs in which all members of society must share. The resources that courts expend in interpreting and refining the ambit of the *sui generis* database right for reasons relating to both legal clarity and public policy are examples. Similarly, the hesitation of prospective users to utilise databases that could be IPR-protected may have a deterrent effect that leads to a systematic decline in the translation of available data into valuable outputs, depriving society of the benefits.

Approaches to reform that would create a registration requirement or specialised administrative body responsible for confirming the ambit and application of the database right alleviate some, but not all of the above difficulties. For example, the operation and decision-making activities of such an administrative organ, and the performance and verification of the associated formalities, still create high costs for database producers, database users, and for society at large. Furthermore, rendering legal rights contingent on performing legal formalities (e.g. registration of the right) can favor well-capitalised commercial entities and academic research institutions, relative to smaller organisations and individuals. The better-capitalised institutions will often retain legal counsel and expend additional resources to ensure that their rights are secured in databases that these institutions generate.

Bearing some of the costs might be worth the while if the *sui generis* database right created benefits that justified these additional costs. However, our analysis demonstrates that the incentives that the database right creates are neither effective in stimulating database production, nor required to such end. Database producers often have incentives to compile databases that are not conditional on the assertion of proprietary rights therein. In those instances in which asserting proprietary rights in databases is valuable, digital rights management or contractual mechanisms are both more effective, and more flexible, than the database right.

There is extensive literature on the potential for legal complexity to arise from the interaction of simple rules. This literature further notes that there are little empirical criteria available to determine whether a particular combination of legal rules is “too complex”, such that it is difficult to measure the complexity of law using a meaningful set of metrics. It is also difficult to compare the success of a law in meeting its objectives, relative to its complexity, or to the complexity it introduces to the legal system.²⁰¹ Despite these limitations, it is possible for us to conclude that the complexity produced by the *sui generis* database right outweighs the benefits that it creates. However, at minimum, in the absence of repeal initiatives, publicly funded databases and those that are made available for downstream use as a matter of law, should be excluded from the operation of the *sui generis* database right.

One final issue that ought to be addressed in reforming the *sui generis* database right is the result of the *Ryanair* decision. In the context of public biological

²⁰¹ Ruhl and Katz (2015); Schuck (1992).

databases, the repeal of the *sui generis* right would require strictures on the use of contracts and DRM technologies to prevent the downstream use of such databases. Indeed, the same critiques that we have levied concerning the transaction costs and prospect for infringement inherent in navigating the *sui generis* database right could be levied at the need to navigate contracts of adhesion, licenses, and DRM technologies that are appended to databases. To this end, we recommend legislating an unqualified right to the reuse of public databases similar to those anticipated in the Database Directive, the Open Data Directive, or the Data Governance Act. This proposal is coherent with the broader EU pivot away from recognising exclusive rights in databases, and toward encouraging the translation of publicly-funded data into valuable downstream outputs.

7 Conclusion: The *Sui Generis* Database Right in Context

To reiterate: the *sui generis* database was devised to foster investment in database production, and to stimulate markets in information and related services. In concise legislation that has remained open to jurisprudential refinement and redefinition, the Database Directive creates a strong and novel IPR, offered to database makers as an incentive to translate pre-existing data into novel databases.

However, the database right has a deleterious effect on the reuse of available data. Firms, public bodies, and research organisations bear strong incentives to generate and compile data for their own purposes, absent the database right acting to incentivise such behavior. Much of the economic value that arises from data use can be ascribed to databases, search engines, and other downstream products that are produced from pre-existing databases. Therefore, the database right does not appear to be needed to ensure that databases are produced. Once recognised in a database, it creates friction with the competing policy imperative to enable the liberal downstream use of data. To promote this latter objective, courts and legislatures have created considerable exceptions to the database right, limiting the circumstances in which it is recognised, the conditions that give rise to its breach, and the circumstances according to which it can be exercised.

These modifications have not led to a streamlined right that elegantly balances incentivising database production against freeing up information reuse. Instead, the database right as now articulated applies in a conditional manner, and determining the circumstances of its application requires jurists to balance numerous competing factors. These include determining whether a database is protected, whether legislation precludes the rights-holder's reliance on the database right, and whether contemplated downstream uses pose a risk to a protected investment.

In performing this evaluation, established firms with considerable resources to allocate to legal counsel are favored relative to smaller organisations. These firms seldom need the protection of the database right, as these hold the resources to use contracts and DRM technologies to protect databases that are valuable to them. Smaller firms, public bodies, and research organisations often conduct their activities through decentralised networks. These organisations will less often have the resources to determine whether their own databases benefit from the database

right, and in fact, the preconditions to the recognition thereof might operate to disfavor such networked production from giving rise to protected databases. In reusing available public databases, it is anticipated that smaller organisations of this nature will refrain from the reuse of available and non-protected databases, for fear of non-compliance with applicable rights. In contrast to larger and more established firms, these do not possess the financial resources and access to in-house legal expertise required to distinguish infringing conduct from non-infringing conduct. The net effect of the database right is therefore to further entrench information monopolies, an outcome that is in direct conflict with the stated objectives of EU information policy. In all cases, establishing whether a database is subject to the database right, and whether a proposed downstream use breaches the database right, imposes high costs on organisations in the EU, leaving them at a competitive disadvantage relative to non-EU organisations that need not direct precious legal compliance resources to performing such analyses.

Our conclusion creates reason to be cautious about the broader EU effort to balance context-specific protected interests in information (e.g. commercial rights, data protection rights) against downstream uses thereof that might conflict with those protected interests. Recent legislative efforts including the General Data Protection Regulation (GDPR), the Data Governance Act, the proposed Data Act, and the proposed European Health Data Space Regulation (EHDS Regulation) attempt to develop concerted public policy proposals directed to balancing individual and/or private interests in information against the framing thereof as part of a commons that is open to unqualified use.

Legal mechanisms used to balance contextually recognised private interests in information against the general public interest in enabling its reuse include individual, enforceable rights, licensing requirements, reporting requirements, and the requirement to describe and to balance risks in impact assessments. Other mechanisms include the creation of novel information regulators, such as supervisory authorities and competent bodies, and of bodies responsible for issuing interpretive guidance, such as the European Data Protection Board (EDPB) and the European Data Innovation Board (EDIB).

The EU effort to invest considerable legislative and public-sector resources in the creation of clear rules balancing contingent rights to restrict information use against the general right to reuse such information is laudable. However, the lesson of the *sui generis* database right is that these efforts will create complicated legal structures that are cost-intensive to understand and to administer. Systematically, reliance on this approach will likely contribute to the indeterminacy of the applicable laws because of the intricacy of the applicable legal tests and the long timespans required to obtain feedback from courts and regulators regarding best efforts compliance. It will also entrench the advantages of centralised organisations that can better coordinate their activities according to a singular and shared response to applicable legislation, and entrench the advantages of large organisations with deep pockets to dedicate to mitigating or withstanding the risk of non-compliance. Last, it will render EU organisations less competitive relative to non-EU organisations acting in the same sphere or activities, as the former are required to bear heightened transaction costs and heightened costs of compliance.

Though suggesting remedies to these broader public policy challenges is outside of the scope of this paper, a few brief comments can be made. First, the iterative effort to review and to refine the real-world effects of EU information policy are welcome. It is difficult to anticipate the effects that legislation will have on the regulated object. Reviewing and recasting legislation can help mitigate its unanticipated deleterious effects, or to achieve similar benefits in a more cost-effective manner. It can also ensure that the law remains responsive to shifting incentives or changing practices in the regulated sector, which motivate a change in regulatory strategy. Second, in favoring the interests of pro-social efforts such as research efforts, bright-line exceptions to the infringement of public law, or to the infringement of private rights, should be favored, instead of contextual and conditional exceptions thereto. This is the case because the organisations engaged in these categories of pro-social enterprise do not have the resources required to engage in intensive legal interpretation, nor to bear the afferent risks. Therefore, a contextually applicable exception to onerous regulatory requirements or to the infringement of private rights, such as the *sui generis* database right, can fail altogether to alleviate the applicable regulatory burden, where the costs of assessing whether such an exemption finds application remain high. These latter approaches – reviewing legislation often, and integrating blanket exceptions favoring privileged categories of actors such as SMEs and research organisations – can help to reconcile the need of low-capacity organisations for low-cost solutions, with the wide and ambitious project of creating comprehensive information policy throughout the EU, through the continuous implementation of sweeping omnibus statutes.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

Journal articles and reports

- Afshinnekoo E, Ahsanuddin S, Mason CE (2016) Globalizing and crowdsourcing biomedical research. *Br Med Bull* 120:27
- Benkler Y (2007) *The wealth of networks: how social production transforms markets and freedom*. Yale University Press, London
- Beunen AC (2007) Protection for databases: the European Database Directive and its effects in the Netherlands, France, and the United Kingdom. Wolf Legal Publishers, Nijmegen
- Borghesi M, Karapapa S (2015) Contractual restrictions on lawful use of information: sole-source databases through the backdoor? *EIPR* 37(8):505–514
- Bottis M (2015) How open data become proprietary in the Court of Justice of the European Union. In: Katsikas SK, Sideridis AB (eds) *E-cemocracy*. CCIS, p 570

- Bui AAT, Van Horn JD, NIH BD2K Centers Consortium (2017) Envisioning the future of big data biomedicine. *J Biomed Inform* 69:115-7
- Buri I (2012) Accessing and licensing government data under open access conditions. Creative Commons Netherlands. Institute for Information Law, University of Amsterdam
- Cardinale PJ (2006–2007) Sui generis database protection: second thoughts in the European Union and what it means for the United States. *Chi-Kent J Intell Prop* 6:157
- Compagnucci MC (2020) Database rights in big data and the cloud—main legal considerations. In: *Big data, databases, and ownership rights in the cloud. Perspectives in law, business, and innovation.* Springer Nature, Singapore
- Contreras JL (2018) The anticommons at 20: concerns for research continue. *Science* 361(6400):335
- Cunha Neves P, Afonso O, Silva D, Sochirca E (2021) The link between intellectual property rights, innovation, and growth: a meta-analysis. *Econ Model* 97:196–209
- Derclaye E (2004) Database sui generis right: should we adopt the spin-off theory? *Eur Intellect Prop Rev* 26(9):402–413
- Derclaye E (2012) Recent French decisions on database protection: towards a more consistent and compliant approach with the Court of Justice's case law? *Eur J Law Technol* 3(2)
- Derclaye E, Husovec M (2021) Sui generis database protection 2.0: judicial and legislative reforms. Unpublished Draft Manuscript
- Desai S (2018) Shhh—it's a secret: a comparison of the United States Defend Trade Secrets Act and European Union Trade Secrets Directive. *Ga J Int Comp Law* 46:2
- Doldirina CA, Eisenstadt H, Onsrud H, Uhler P (2018) Legal approaches for open access to research data. <https://doi.org/10.31228/osf.io/n7gfa>
- Drahos P, Braithwaite J (2002) *Information feudalism: who owns the knowledge economy?* Routledge, Milton Park
- Dursi L et al (2021) CanDIG: federated network across Canada for multi-omic and health data discovery and analysis. *Cell Genom* 1(2):100033
- Dyke SOM (2020) Genomic data access policy models. In: *Responsible Genomic Data Sharing.* Academic Press pp. 19–32
- Eisenberg RS (2008) Noncompliance, nonenforcement, nonproblem? Rethinking the anticommons in biomedical research. *Hous L Rev* 45(4):1059
- European Commission (2005) DG internal market and services working paper: first evaluation of Directive 96/9/EC on the legal protection of databases
- European Commission (2018) Study in support of the evaluation of Directive 96/9/EC on the legal protection of databases
- European Commission (2022) Impact assessment report accompanying the proposal for a regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act)
- Fecher B, Friesike S (2014) Open science: one term, five schools of thought. In: Bartling, Friesike (eds) *Opening science.* <https://doi.org/10.1007/978-3-319-00026-8>
- Freeberg MA et al (2022) The European genome-phenome archive in 2021. *Nucleic Acids Res* 50(D1):D980–D987
- Greshake B, Bayer PE, Rausch H, Reda J (2014) OpenSNP – a crowdsourced web resource for personal genomics. *PLoS ONE* 9(3):e89204
- Hadfield GK (1988) The economics of copyright: an historical perspective. *Copyright Law Symp* 38:1
- Heymann LA (2008) Everything is transformative: fair use and reader response. *Colum J Law Arts* 31(4):445
- Huarrag Guerrero EM (2017) Challenges of intellectual property in the information society. W.I.P.O.—W.T.O. Colloquium Papers
- Hugenholtz PB (2000) Implementing the European Database Directive. *Int Intell Prop Law Policy* 70(4):1
- Hugenholtz PB (2001) The new database right: early case law from Europe. *Fordham University School of Law, Ninth Annual Conference on International I.P. Law and Policy*
- Hugenholtz PB (2016) Something completely different: Europe's sui generis database right. In: Frankel S, Gervais D (eds) *The internet and the emerging importance of new forms of intellectual property.* (Information Law Series; vol. 37). Wolters Kluwer
- Keller P (2022) A vanishing right? The sui generis database right and the proposed Data Act. *Kluwer Copyright Blog*
- Koo A (2010) Database right decoded. *Eur Intellect Property Law Rev* 32(7):313–319

- Koščík M, Myška M (2017) Database authorship and ownership of sui generis database rights in data-driven research. *Int Rev Law Comput Technol* 31(1):43–67. <https://doi.org/10.1080/13600869.2017.1275119>
- Lappalainen I et al (2015) The European genome-phenome archive of human data consented for biomedical research. *Nat Genet* 47(7):692–695
- Lee P (2008) The evolution of intellectual infrastructure. *Wash Law Rev* 83:1
- Liu J (2019) An empirical study of transformative use in copyright law. *Stanf Technol Law Rev* 22:1–163
- Lipton JD (2009) Wikipedia and the European Union Database Directive. *Santa Clara Comput High Technol Law J* 26(4):631
- Maccari G, Robinson J, Hammond JA, Marsh SE (2020) The IPD Project: a centralised resource for the study of polymorphism in genes of the immune system. *Immunogenetics* 72:49–55
- Malsagova K et al (2020) Biobanks: a platform for scientific and biomedical research. *Diagnostics* 10:485
- Margoni T, Kretschmer M (2022) A deeper look into the EU text and data mining exceptions: harmonisation, data ownership, and the future of technology. *GRUR Int* 71(8):685–701
- Myska M, Harasta J (2016) Less is more: protecting databases in the EU after Ryanair. *Masaryk Univ J Law Technol* 10(2):170
- O’Doherty KC et al (2021) toward better governance of human genomic data. *Nat Genet* 53:2–8
- Olk P, West J (2020) The relationship of industry structure to open innovation: cooperative value creation in pharmaceutical consortia. *Res Dev Manag* 50(1):116–135
- Posner RA (2005) Intellectual property: the law and economics approach. *J Econ Perspect* 19(2):57–73
- Poupon V, Seyller A, Rouleau GA (2017) The Tanenbaum Open Science Institute: leading a paradigm shift at the Montreal Neurological Institute. *Neuron* 95:1002
- Rai AK (2017) Risk regulation and innovation: the case of rights-encumbered biomedical data silos. *Notre Dame Law Rev* 92:1641
- Rambla J et al (2022) Beacon v2 and beacon networks: a “lingua franca” for federated data discovery in biomedical genomics, and beyond. *Hum Mutat* 43(6):791–799
- Ross JS, Lehman R, Gross CP (2012) The importance of clinical trial data sharing: toward more open science. *Circ Cardiovasc Qual Outcomes* 5(2):238–240
- Ruhl JB, Katz DM (2015) Measuring, monitoring, and managing legal complexity. *Iowa Law Rev* 101(1):191
- Saez-Rodriguez J et al (2016) Crowdsourcing biomedical research: leveraging communities as innovation engines. *Nat Rev Genet* 17:470
- Schellekens M (2011) A database right in search results? *Comput Law Secur Rev* 27:620–629
- Schuck PH (1992) Legal complexity: some causes, consequences, and cures. *Duke Law J* 42(1):1–52
- Simões de Almeida G, Faria Abreu G (2021) Crowdsourced databases and *sui generis* database rights. [arXiv:2108.04727](https://arxiv.org/abs/2108.04727)
- Stamatoudi IR (1997) EU Database Directive: reconceptualising copyright and tracing the future of the sui generis right. *RHDI* 50:435
- Sudlow C et al (2015) UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 12(3):e1001779
- The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium (2020) Pan-cancer analysis of whole genomes. *Nature* 578:82–93. <https://doi.org/10.1038/s41586-020-1969-6>
- Tryka et al (2014) NCBI’s database of genotypes and phenotypes: dbGaP. *Nucleic Acids Res* 42:D975–D979
- United Nations Economic, Social, and Cultural Organisation (UNESCO) (2021) UNESCO Recommendation on Open Science
- University of Amsterdam, Institute for Information Law (2006) Database right case law collection citing: *Spot (cinebel.be) v. Canal Numédia (allocine.be)*. District Court (Tribunal de première instance) Brussels, 18 January 2002
- Varian HR (1998) Markets for information goods. University of California, Berkeley
- World Intellectual Property Organisation (WIPO) (2002) Standing Committee on Existing Legislation Concerning Intellectual Property in Non-Original Databases. Eighth Session, Geneva, November 4 to 8, 2002. “Summary on Existing Legislation Concerning Intellectual Property in Non-Original Databases” SCCR/8/3
- Zhang et al. (2011) International cancer genome consortium data portal – a one-stop shop for cancer genomics data. *Database*

Jurisprudence

- ABRvS 29 April 2009, n 07/786, AMI 2009-6 (College B&W Amsterdam/Landmark; m. nt. M. Van Eechoud)
- Case C-762/19, *CV-Online Latvia SIA v. Melons SIA* ECLI:EU:C:2021
- Case C-304/07 *Directmedia Publishing GmbH v. Albert-Ludwigs-Universität Freiburg* ECLI:EU:C:2008:552
- Case C-202/12 *Innoweb B.V. v. Wegener I.C.T. Media B.V. and Wegener Mediaventions BV* ECLI:EU:C:2013:850
- Case C-203/02 *The British Horseracing Board Ltd and Others v. William Hill Organization Ltd.* ECLI:EU:C:2004:695 (hereinafter British Horseracing Board)
- Case C-46/02 *Fixtures Marketing Ltd v. Oy Veikkaus Ab* ECLI:EU:C:2004:694
- Case C-444/02—*Fixtures Marketing Fixtures Marketing Ltd v. Organismos prognostikon agonon podosfairou AE (OPAP)* ECLI:EU:C:2004:697
- C-338/02 *Fixtures Marketing Ltd v. Svenska Spel AB* ECLI:EU:C:2004:696
- Case C-30/14 *Ryanair Ltd v. PR Aviation BV* ECLI:EU:C:2015: 10

Legislation

- Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, Official Journal of the European Communities, L 77/20
- Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society, OJ L 167, 22.6.2001(Information Society Directive)
- Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the Re-Use of Public Sector Information. Official Journal of the European Union L 345/90 (PSI Directive)
- Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use, and disclosure (Trade Secrets Directive)
- Directive 2019/1024 of the European Parliament and of the Council of 20 June 2019 on Open Data and the Re-Use of Public Sector Information (Recast) L 172/56 (Open Data Directive)
- Proposal for a Council Directive on the legal protection of databases, COM(92) 24 final—SYN 393, (Brussels, May 13, 1992); Explanatory Memorandum
- Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act) COM/2022/68 final
- Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC OJ L 119, 4.5.2016 (General Data Protection Regulation)
- Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 PE/85/2021/REV/1OJ L 152, 3.6.2022, pp. 1–44 (Data Governance Act)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.