Deliverable No. D4.4

Whitepaper -

Recommendations for an amended European legal framework on patients' and researchers' rights and duties in E-health related research

Grant Agreement No.:	600841
Deliverable No.:	D4.4
Deliverable Name:	Whitepaper - Recommendations for an amended European legal framework on patients' and researchers' rights and duties in E-health related research
Contractual Submission Date:	31/03/2016
Actual Submission Date:	31/03/2016

Dissemination Level		
PU	Public	**X**
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

COVER AND CONTROL PAGE OF DOCUMENT	
Project Acronym:	**CHIC**
Project Full Name:	Computational Horizons In Cancer (CHIC): Developing Meta- and Hyper-Multiscale Models and Repositories for In Silico Oncology
Deliverable No.:	D4.4
Document name:	Whitepaper - Recommendations for an amended European legal framework on patients' and researchers' rights and duties in E-health related research
Nature (R, P, D, O)[1]	R
Dissemination Level (PU, PP, RE, CO)[2]	PU
Version:	1.0
Actual Submission Date:	31/03/2016
Editor: Institution: E-Mail:	Nikolaus Forgó Leibniz Universität Hannover nikolaus.forgo@iri.uni-hannover.de

[1] **R**=Report, **P**=Prototype, **D**=Demonstrator, **O**=Other

[2] **PU**=Public, **PP**=Restricted to other programme participants (including the Commission Services), **RE**=Restricted to a group specified by the consortium (including the Commission Services), **CO**=Confidential, only for members of the consortium (including the Commission Services)

ABSTRACT:

This deliverable D4.4 is a report on the European legal framework on patients' and researchers' rights and duties in E-health-related research and aims at giving recommendations to the European Commission and other political stakeholders on the ongoing reforms in the European legal framework on the relevant subject areas, including data protection, medical devices rules and IPR.

KEYWORD LIST:

Data Protection, In silico models, Medical Devices, Clinical Investigations, Copyright

MODIFICATION CONTROL			
Version	**Date**	**Status**	**Author**
0.1	09/12/2015	Outline Draft	Iheanyi Nwankwo
0.2	12/02/2016	Draft	Alan Dahi
0.3	20/02/2016	Draft	Marc Stauch
0.4	26/02/2016	Draft	Alan Dahi; Iheanyi Nwankwo
0.5	03/03/2016	Updated Draft	Marc Stauch
0.6	07/03/2016	Revised Draft	Iryna Lishchuk
0.7	17/03/2016	Revised Draft	Marc Stauch
0.8	19/03/2016	Prefinal Draft	Nikolaus Forgó

List of contributors

- Alan Dahi, LUH
- Nikolaus Forgó, LUH
- Iryna Lishchuk, LUH
- Iheanyi Nwankwo, LUH
- Marc Stauch, LUH

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n^{o} 600841.

Contents

1 Executive Summary

This whitepaper is a report on the European legal framework on patients' and researchers' rights and duties in E-health-related research and aims at giving recommendations to the European Commission and other political stakeholders on the ongoing reforms in the European legal framework on the subject matter. Although E-health describes the application of ICTs across the whole range of functions in the healthcare sector, this whitepaper will focus on medical research activities that utilize the application of information and communication technology (ICT) for the processing and use of (patient) data to gain scientific knowledge. Particular attention will be paid on the amalgamation of models in the field of computer-assisted medical research – *in silico*-based medical research.

The report is divided into three main parts. The first part (Chapter 3) centers on the general framework in which medical research has been conducted in the modern era (post-WWII), and challenges to the framework posed by the ermergence of new modes of E-health research. As will be described, the challenge is not directed to the substance of what the established principles of good research practice seek to achieve (in terms of the high value they place upon the integrity and autonomy of the research subject), but whether as drafted they are still apt in the novel and innovative context of E-health-related research to achieve those ends, and without imposing unnecessary restraints upon researchers to pursue important research in the public good.

In the second part (Chapters 4 and 5), we look in detail at two central sets of legal rules in this regard, which have been the subject of recent reform activity at EU level, namely data protection law and the rules for validating and certifying medical devices (including E-health-related support tools). We also seek to anticipate future regulatory challenges, should the use of such tools, and in particular predictive models, become widespread. In the third part (Chapter 6), we then evaluate how well current intellectual property rights terms facilitate reward for researcher efforts in developing innovative E-health approaches, such as in silico cancer models, to assist clinical decision-making, and will suggest ways of improving the current practice. The report then concludes (Chapter 7) by giving concrete recommendations to the Commission and other stakeholders for reforms in this area.

2 Introduction

The global healthcare system is struggling with rising costs and uneven quality despite various policies aimed at improving the system. A particular challenge is the management of chronic diseases of an unpredictable nature, such as cancer or Parkinson's disease, that affect each individual differently and progress in diverse ways. Here, the use of information technologies in tackling this problem have shown some positive outcomes; advances in health IT including the use of powerful cloud computing have enabled various transformations in translational medicine, ranging from genomic sequencing to the availability of large bioinformatics databases. Further research into these forms of health care research and delivery may be regarded as a political and economic necessity: according to the EU's eHealth Report, demographic change in Western countries is rendering the costs of their traditional health care systems unsustainable.[3]

Achieving more efficient health care, by targeting treatments to the patients best able to benefit from the specific therapy, will require ICT decision-support based on as much relevant data as possible.[4] But at the same time (and of great ethical importance), the individual patient too stands to benefit from better treatment, in addition to avoiding unnecessary, ineffective or harmful treatment. With these advancements, it is now possible to integrate clinical and molecular sciences with computing technologies for treatment and research purposes.

These advances have also meant changes in the technical circumstances in which law that regulates these activities operate, including in medical research in which a computer-driven revolution has taken place with the ultimate goal of achieving personalized treatment by better understanding the individual reasons for the disease. This requires a lot of computing and a lot of division of work and, therefore, a lot of data transfer for research as well as for treatment reasons.[5] Nowadays, we see various health-related applications and devices in the hands of the ordinary citizen or patients which generate data with potential value for clinical care, research, and other exploitations. Due to the ongoing digitalisation of the medical field and the increase of readily accessible available data, medical research questions can now often be investigated outside of traditional interventional clinical trials. This development has been recognised by the European Commission, with. E-health a major focus of its Digital Agenda for Europe.[6] A lot of research would benefit from the secondary use of medical data or the use of ICT for observational studies.

[3] European Commission 'Redesigning health in Europe for 2020', pp. 7-8; see also the European Commission Green Paper on mobile health ("mHealth"), COM(2014) 219 final, p. 13. Retrieved 27 September 2015, ec.europa.eu/digital-agenda/en/news/green-paper-mobile-health-mhealth.

[4] 'Redesigning health in Europe for 2020', Section 1.

[5] Nikolaus Forgó, 'My health data—your research: some preliminary thoughts on different values in the General Data Protection Regulation'.

[6] European Commission, https://ec.europa.eu/digital-agenda/en/eu-policy-ehealth. See also COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS eHealth Action Plan 2012-2020 - Innovative healthcare for the 21st century, COM/2012/0736, http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52012DC0736.

To this end, a number of ICT infrastructure elements that facilitate data management, access, sharing, analytics and other processes in the health sectors have been developed in the last decades, also facilitating medical research, and more and more energy is being put to work in order to translate the results from research to the bedside. For instance, the development of systems biology has only been possible through the application ICT to handle the large volume and variety of data about molecular processes in cells and organisms.[7] And the knowledge gained from these processes will be used in clinical care.

The Patients' Rights in Cross-Border Healthcare Directive[8] (Directive 2011/24/EU) aims to provide "*rules for facilitating the access to safe and high-quality cross-border healthcare and promotes cooperation on healthcare between Member States*".[9] It also provides for the establishment of voluntary "European reference networks",[10] (ERNs) which shall support cooperation between healthcare providers and centers of expertise, for example to strengthen research.[11] In addition, Article 14 encourages the development of voluntary eHealth networks,[12] with one of their goals being guidelines on "*effective methods for enabling the use of medical information for public health and research*".

The ERNs envisaged by Directive 2011/24/EU could serve as a model to facilitate cross-border E-health research. The National Action League for People with Rare Diseases is an example of a network with multiple (domestic) stakeholders which could become part of an ERN. Its National Plan of Action[13] also focuses on the action field of research with the goals of improving the overall conditions for research and development in the area of rare diseases, accelerating the transfer of research knowledge to practical applications, improving research on optimizing therapy, and improving health services research. So-called reference centers in the network are to "provide a multidisciplinary research infrastructure where both basic and clinical research as well as research into healthcare provision can be carried out",[14] together with centers of expertise and cooperating centers.

A number of projects have also been funded over the years by the EU in this area,[15] and while the ultimate aim of translating these research outcomes into daily clinical practice is still fully to be met, the chances are high that we will see these translations in the near future. However, a number of challenges are still facing this area of research – technological and legal. As we will show in this report, understanding how these facilities interact with data processing, research, and medical device regulations is profound in framing the right

[7] I. Peterson, et al., Systems Biology, Information Technology and Cancer Research in M. Döring et al., Contextualizing Systems Biology, Springer, Switzerland, 2015, p. 147.

[8] Directive 2011/24/EU of the European Parliament and of the Council of 9 March 2011 on the application of patients' rights in cross-border healthcare.

[9] Article 1 (1) 2011/24/EU.

[10] Article 12 2011/24/EU.

[11] Article 12 (1), (2)(e) 2011/24/EU.

[12] Article 42 2011/24/EU.

[13] National Action League for People with Rare Diseases, National Action Plan http://www.namse.de/images/stories/Dokumente/Aktionsplan/national%20plan%20of%20action.pdf.

[14] Ibid., p. 10.

[15] See J. Herveg et al, Final Recommendations on Legal Issues in Ehealth (Legally eHealth project D.5. (PUB)), at p 3, referring inter alia to the FP6 Eurogentest and Lifescihealth projects; see also the FP7 projects , VPH Share [www.vph-share.eu] and MyHealthAvatar [www.myhealthavatar.eu].

policy for E-health research and development. This, in essence, means untangling the tension between “old laws and new techniques” as the cases are increasing in which rules need to be applied to circumstances they were not intended to cover.[16]

In the subsequent chapters we look in more detail at the ethical and legal challenges posed for patient and researcher rights in relation to E-health related medical research. As will be seen, the European health framework has already been subject to significant recent reform, with new EU secondary legislation introduced or pending in the areas of clinical drug research, data protection, and medical devices certification; however, it remains open to discussion how far these changes will assist research in the specific field of E-health-related VPH and in silico modelling. A key challenges here arises from the need to accommodate the rights of patients/research subjects to the new modalities of research, which poses difficulties to the mechanism of consent and exposes patients to potential new risks of harm and distress. In this report we will use the case of the in silico-based research designed for the CHIC project as an example in elucidating this relationship, and pinpointing areas where reform in the regulatory framework is desirable in order to help advance and foster E-health innovations.

[16] Forgó N, et al.(2010): Ethical and Legal Requirements for Transnational Genetic Research, München: C.H.Beck Verlag.

3 E-Health including In Silico Research, and Challenges for the Traditional Framework of Medical Research

3.1 Introduction

As a sub-aspect of E-health-related research , in silico-based medicine describes the modeling, simulation, and visualization of biological and medical processes in computers.[17] A key aim of this area of medical research is to develop models representing different aspects of the human biological system, thus allowing biological processes to be simulated and studied virtually rather than in a laboratory (in vitro) or in the live subject (in vivo). Systems biology, for example, aims to model molecules, cells, tissue, organs, body systems, and whole organisms holistically, and it is believed that one possible way to reach this goal is by studying diseases and the alterations between normal and diseased biospecimens, and theoretically analyzing them with in silico techniques.[18] In in silico oncology, for instance, the basic idea is that, when a cancer patient comes in for treatment, the "oncosimulator" canl be used to predict what the tumor will do when treatment is given, and which treatment will produce the best outcome - for example, radiation or chemotherapy.[19]

The fundamental assumption underlying in silico-based research is that any biological process is amenable to mathematic and/or algorithmic description. In this regard, the genesis and development of cancer are regarded as a disease and at the same time as a natural phenomenon that may be understood scientifically. Thus, subsystems of the tumor can be studied and a reproducible model of a specific cancer and its progression built.[20] To this end, the knowledge captured in models at different scales of the biological process (molecular, cellular, tissue, organ, etc) will be integrated into composite models (hypermodels) of increasing complexity, capable of simulating processes in relation to a given disease domain.

The ultimate aim would be to fuse hypomodels covering discrete domains into an overall linked-up hypermodel covering the human patient as a whole. The latter could then be fed with data relating to an individual patient in order to answer specific disease treatment questions in relation to that patient. The whole system could be used to create a digital patient, i.e. a virtual version (an avatar) of each living person, as well as to run simulations of health and disease processes on this virtual individual . Thereafter, outputs from the model could be used to make not only decisions in response to disease but long range predictions about that individual's future health, allowing preventative strategies and interventions to be employed before diseases appear.

Apart from the above, the construction or design of an *in silico* clinical trial could provide profound insight into the design of real life clinical trials, ranging from optimal patient selection to individualized dosage and duration of proposed therapeutic interventions. There may well be other benefits of such prior checking *in silico*, which include the targeting of

[17] Peterson et al, Systems Biology, Information Technology and Cancer Research.

[18] See Coveney et al, Computational Biomedicine.

[19] Ibid, p 180.

[20] Ibid., p 174.

drugs based on individual patient profiling, reduced animal testing, identifying problematic side effects, creating tailored treatments, understanding costs and benefits at an individual level, etc.[21] Trials can be run harmlessly on individual digital human models, as well as on entire virtual patient populations numbering hundreds or thousands, and which may reduce the need for subsequent *in vivo* tests. Where this trend is sustained, *in silico* clinical studies will reduce cost and error to a great extent, and aid in achieving the goals of personalized medicine.

The potential reduction of the risk of causing physical harm compared to *in vivo* research on real patients and volunteers is a good ethical reason to switch to *in silico* research, so far as the reliability of the results is acceptable. Elsewhere, *in silico* medicine has the potential to generate further positive benefits by unlocking new knowledge from patient data that can be used to optimize individual care and treatment. There have been some success stories in this area such as the virtual self-surviving cell modeled by Masaru Tomita,[22] to predict the likelihood of chemotherapy benefit for patients with low-grade breast cancer, and in the prediction of tumor shrinkage in nephroblastoma.[23] Further, there is the potential that a construction or design of an *in silico* clinical trial could be used to provide profound insight into the design of real-life clinical trials, ranging from optimal patient selection to individualized dosage and duration of proposed therapeutic interventions.[24]

Some feedback from clinicians already indicate the benefits of using model simulations for better diagnoses and treating patients.[25] Thierry Marchal, though, has noted two important entry barriers preventing a larger adoption of simulation in the clinical world - complexity and time;[26] we would add further that legal issues surrounding building the models, validating them, publication of outcome and intellectual property rights also pose challenges. It is submitted that legal and ethical landscape upon which medical research is generally conducted is largely in the context of the traditional model (directly using human subjects), where rules were made primarily to protect the human research subjects from harm. Similarly, IP rights to reward research investment were premised on the production of tangible devices and drugs that fell comfortably within the traditional concepts of patent and other IPR regimes. In this regard, the question now is: should E-health research still be subject to the current legal and ethical rules or should new ones be developed to cater for the needs of the *in silico* community and facilitate their work?

[21] See: http://avicenna-isct.org/projectinformation.html

[22] Tomita, M (2001) 'Whole-cell simulation: a grand challenge of the 21st century'.

[23] Stamatakos G, et al. (2014), 'The technologically integrated oncosimulator: combining multiscale cancer modeling with information technology in the in silico oncology context'.

[24] Coveney et al, Computational Biomedicine, p 182.

[25] https://www.linkedin.com/pulse/simulation-every-clinician-thierry-marchal?trk=v-feed

[26] Ibid.

3.2 The traditional medical research model

As discussed, the legal and ethical landscape upon which medical research is conducted is premised largely on the traditional model (directly using human subjects), where rules were made primarily to protect the human research subjects from harm. Following major atrocities committed against patients and research subjects during the Nazi era, as revealed in the Nuremberg trials, and other unethical practices that have occurred in the past, such as exploitation of subjects in the Tuskegee study of syphilis, the issue of patients' right during medical research has received a lot of attention.[27] The Nuremberg Code of 1947 was formulated to protect patients' interest during medical research and states that "the voluntary consent of the human subject is absolutely essential" before conducting such research.[28] The Declaration of Helsinki, adopted by the WMA in 1964 and updated on several occasions since, has elaborated on the principles of medical ethics to be adhered to during research.[29]

Broadly speaking, three main principles may be said to underlie the approach to research practice involving human subjects set out in the Declaration. First, as already mentioned, the subject's voluntary informed consent, untainted by coercion and based on an appreciation of the risks posed by the research, is essential.[30] Secondly, the researcher must in any event take all appropriate steps to minimize risks to the subject; the latter must be proportionate to the hoped for knowledge to be gained by the study. For this reason (as we shall consider below in relation to clinical trials) it is generally required for new medicinal products to divide trials into phases, first testing the danger of adverse toxic side effects in small groups of volunteers (phase I) before testing the drug's positive efficacy in larger groups of patient/subjects (phase II and III). Importantly, too, the researcher's duty also extends to minimizing privacy risks by upholding the subject's confidentiality.[31]

Thirdly, the research must be based on sound science and be carefully planned to produce meaningful results that have a reasonable chance of being replicated in clinical practice. In this regard, the 'randomised control trial' (RCT) is generally seen as the scientific gold standard for validating such research. The scientific utility of randomisation is derived from the elimination, through the randomisation process, of potential bias in observing and interpreting the results of the treatment. Thus, in the context of pharmaceutical trials, some participants will not receive the new drug under study, but will randomised into a 'control' group that receives the existing drug therapy, or – if there is none, a placebo (i.e. an inert substance, such as a 'dummy' pill, which is given so the subject is unaware that he is in the control group not receiving the drug under study). This is meant to deal with the beneficial effects of a patient simply believing they are receiving a genuine therapy designed to make them better.

Since the said 'placebo effect' can also influence the interpretation of trial outcomes by the researcher, a lot of clinical trials take place 'double blind' so the researcher similarly does not know which subjects are receiving the new drug, as opposed to in the control group. The

[27] Pedroni J, Pimple K (2001): A Brief Introduction to Informed Consent in Research with Human Subjects.

[28] [http://www.cirp.org/library/ethics/nuremberg/].

[29] See the current (2013) DoH version, at: [http://www.wma.net/en/20activities/10ethics/10helsinki/].

[30] There are specific rules dealing with research on mentally incapacitated subjects, not considered here.

[31] See DoH, Principle 24.

use of RCTs has been well established over many years as a good research tool, and there are many sophisticated variants on the basic new drug/old drug or placeo two-group model; at the same time, other treatment-research may be less easy to design in this way. Thus, while there are past instances of research being carried out involving placebo surgery, where the patient undergoes some form of surgical intervention (cutting open the body and suturing it without doing anything further) this is now widely regarded as unethical. In particular, whereas with the use of placebo drugs, the subject will at most be denied a possible benefit from an active drug, with surgery (even under local anaesthetic only) there is invariably some active harm to the subject.[32]

In the fourth place, and as an essential adjunct for ensuring the three key principles above are observed in practice, there is a system of independent review, in which the research investigator is required to submit the plan for the research (known as the clinical trial protocol) to an ethics committee or institutional review board for it to assess if the above principles have been satisfied before granting approval.[33] Issues considered during ethical review of scientific projects typically relate to matters such as the presence of fully informed consent, the level of risk to the participants that is acceptable in the light of potential benefits, and the amount of protection afforded to the identities and personal information of participants.

The principles enshrined in the Declaration of Helsinki are echoed and further concretised in numerous codes of good research practice, at international, regional and local level. Internationally, the most important source of soft-law good practice in pharmaceutical drug trials is the International Conference on Harmonisation of Pharmaceutical Trials, Good Practice Guidelines.[34] In the EU, pharmaceutical research is subject to the legal norms set out in the Clinical Trials Directive (2001/20/EC), which is due to be replaced by the Clinical Trials Regulation (Regulation No. 536/2014), when this enters force on 28 May 2016. Besides the need for protecting the physical integrity of research subjects, the Regulation also addresses the use of clinical research data in some detail. As this aspect is, as we have noted, especially salient for E-health-related research, including in silico modeling, we consider its key provisions in this regard below

[32] Since failure to benefit can also be seen as a form of harm, the use of placebo drugs should also occur against the background of reasonable uncertainty as to the active drug's beneficial effects. The same goes for the (frequent) case where new therapy is tested against the best existing therapy, rather than a placebo as such.

[33] See DoH, Principle 23.

[34] Available at [http://ichgcp.net/].

3.3 The Clinical Trials Regulation

The Clinical Trials Regulation (CTR), which is due to enter force on 28 May 2016, will apply to all clinical trials conducted in the European Union.[35] A clinical trial is understood as a scientific investigation into the effects of a medicinal product on humans.[36] The CTR mentions the data generated by clinical trials already in Recital 1 and requires it to be reliable and robust, which reflects the data processing principle of "accuracy"[37] . This is further underlined in Recital 17, which requires that the "*authorisation to conduct a clinical trial should address all aspects of subject protection and data reliability and robustness*". At the same time, the term "data" here also refers to non-personal data, ie general scientific data about the medicinal product being investigated – an understanding supported by, inter alia, Recitals 11 and 25, which refer to the data being published in scientific journals and databases, respectively.

As regards the processing of trial subjects' personal data, Recital 29 refers to the "*applicable law on data protection*", and suggests that research institutions should "*be able to collect data from clinical trials to be used for future scientific purposes*", albeit with the subject's consent for the use of the data outside the clinical trial. This informed consent should generally be given in writing.[38] Recital 76 further states, with regard to the withdrawal of informed consent to participate by the research subject, that such a withdrawal should not affect the results of activities, such as the use of the data, carried out before the withdrawal. Finally, Recital 83 reiterates the need for the CTR to be applied in accordance with, inter alia, the principle of respect for private life.

The Articles of the CTR restate the principles explained in the Recitals. For example, Article 28 (1) (d) CTR also guarantees the subjects' rights to privacy and protection of personal data within the clinical trial, and Article 28 (2) CTR opens up the possibility for the use of personal data outside the trial, subject to consent and the applicable data protection law. Article 56 CTR is the main provision on data protection in the Regulation. It states that all personal data shall remain protected in accordance with general data protection legislation and that both technical and organisational measures shall be implemented to do so. Article 93 CTR explicitly mentions the applicability of the EU data protection rules to the processing of personal data carried out pursuant to the CTR. With regard to the reporting obligations of the sponsor of clinical trials, Article 43 (3) CTR states that the annual report on the safety of the investigational medicinal products shall contain only aggregate and anonymised data.

Because of the reference to the applicable data protection legislation, the main problems with the CTR in relation to the processing of personal data for further scientific purposes mirror those that exist in that legislation. Specifically, because the CTR demands that the collection of personal data for future scientific research should be possible with the data subject's consent, the crux point is that of consent. The Data Protection Directive 95/46/EC (DPD) currently places very high requirements on consent to be valid – broad consent is generally not seen as sufficient, since it is seen as neither specific nor informed enough. This leads to difficulty where a data subject does not exactly know in detail what type of research

[35] Article 1 Clinical Trials Regulation.

[36] Article 2 (1), (2), and (3) Clinical Trials Regulation.

[37] See further, section 4.2. below.

[38] Recital 30 Clinical Trials Regulation.

she is consenting to, but only has a broad idea (simplified: “cancer research” vs “onsite research on whether genes X and Y increase the likelihood for developing breast cancer in Caucasian females”). As discussed below in Chapter 4, though, this may change in the future with the enactment of the General Data Protection Regulation (GDPR), which appears more open to the possibility of broad consent.

3.4 Key regulatory challenges

As noted, in silico research in some respects carries significant ethical advantages over traditional medical research. Most obviously, there is no direct risk of physical harm to the subject during the research itself. As in silico research relies heavily on patients’ data, this implies that the rules governing the processing of sensitive data will apply in a number of scenarios – during the development of the models as well as their interaction with other patient data (such as in EHRs). This brings to the fore the privacy and data protection issues that may arise in this aspect of medicine, including during the validation, adaptation, and optimization of the simulations, at which stage the real response of the patient to the treatment is compared with the predicted response.

From a legal perspective, these data are managed and stored in separate databases (in some cases in different geographical sites) by different entities or data controllers, which raises ethical and legal issues relating to confidentiality, data protection and data security.[39] For example, one of the legal and ethical basis for transferring these data into a research environment is the consent of the data subjects, which under the current rules is in most cases understood to be an explicit consent for the use of the data for specific purposes. However, this may raise practical or scientific problems. Frequently, the data at issue will be retrospective, perhaps collected years before, and contacting patients to agree to the new use will be an immensely difficult or even impossible (or unethical) task.[40]

Further risks may arise at the subsequent stages of validating the models (where the directly identifiable data of patients on whom the models are tried will be required, and in the subsequent deployment of models in clinical decision-making. These include potential data breaches, risks to subject autonomy (including through new possibilities to individualise health risks in the insurance market), as well as risks of other harm, such as potentially when a model prediction proves inaccurate and results in a suboptimal treatment outcome. On the other side, though, it is also important to consider the position of the in silico researcher and how far current regulation provides a sufficiently certain field of action, both in terms of avoiding disproportionate or unclear liability risks, and in conferring adequate reward incentives (through the mechanism of IP). We shall look further at these different risks and interests in turn.

3.4.1 Patient/Subject rights and interests

As regards implications for patient rights and interests that arise during the building and integration of hypermodels, the key issues relate to patient privacy and data protection.

[39] Forgó et al, note 17 above.

[40] Lowrance, W, Learning from Experience: Privacy and the Secondary Use of Data in Health Research, 2002, at p 21 ff.

Frequently, this data will be retrospective, raising the issue of consent as discussed above, and the secondary use of clinical and other health-related data. Furthermore, for validation, adaptation, and optimization of models, data is also required for comparing the real response of the patient to the treatment with the predicted response. Here both retrospective and prospective data will be necessary, and the more patients used for validation, the better for the predictive potential of a particular set of models or hypermodels through refinement from feedbacks.

Besides the need to minimise harms to users associated with the unauthorised use of their data, it will clearly be important in ethical and legal terms, and to justify the costs to the health care system of adoption, that VPH / in silico models work safely in delivering accurate and reliable health information to treating clinicians. Here, as in the area of data protection, it will be necessary to think carefully about the risks such systems create in the context of their concrete use. In this regard, the clearest risk arguably stems from the provision of wrong or misleading clinical advice and the harmful consequences of inappropriate therapy that may follow.

These issues, and how they are addressed by current legal regulation, will be explored in detail in Chapter 4 (for data protection) and Chapter 5 (for medical device regulation and potential ex post facto liability), respectively.

3.4.2 Researcher rights and interests

Article 13 of the Charter of Fundamental Rights recognises freedom of research in the EU. This freedom has been elaborated in the European Charter for Researchers[41] and the Code of Conduct for the Recruitment of Researchers,[42] where the EU defines the roles and responsibility of researchers. Although the European Charter and Code of Conduct deal mainly with the relationship between researchers and employers and funders in both public and privacy sectors, they recognize that certain limitations could be imposed on the researchers based on circumstances such as IP protection. Rights of research subjects as well could impose some limitations as to what researchers could do with the research subject. For instance, in medical research that involve intervention to the human body, we saw that the DoH, as well as other international, regional and local rules, imposes an obligation on the researcher to obtain the consent of the research subjects or their legal guardian. Further, there is obligation not to harm the research subjects and to explain the procedure of the intervention before carrying it out. Similarly, for non-interventional research involving only patient data, data protection rules also creates obligations on the researcher, who is seen in this case as the data controller or processors. Here, data protection rules influence the collection, processing and disclosure of health-related data.

Naturally, the fact that rights and interests of research subjects operate as a constraint on medical research is not something researchers are troubled by (and far less, entitled to complain about) per se. To the contrary, researchers – especially where they are medical professionals – will generally be concerned to satisfy the highest medical ethical standards. Rather, the problem is one of normative / legal uncertainty: the unclear scope of rules that

[41] http://ec.europa.eu/euraxess/index.cfm/rights/europeanCharter

[42] http://ec.europa.eu/euraxess/index.cfm/rights/codeOfConduct

purport to protect subject rights, and difficulties applying them in daily research practice. As noted earlier, there is particular challenge in interpreting what the rules require in the context of innovative methods of research, involving health data mining and ICT, which the rule-maker likely did not have in mind. While, in the context of individual research projects, it has been possible to develop secure internal data-sharing frameworks - such as that in CHIC – in which partners access and use securely de-identified data in accordance with contractually stipulated duties of care,[43] the general climate for obtaining and utilising potentially valuable patient data remains challenging, with a tendency towards highly-cautious and diverse access policies.[44] These issues may have a significant impact on researcher freedom, both in the form of ex ante bureaucracy and delay (e.g. the need for multiple ethics committee approval applications that then produce inconsistent responses), and the uncertain liability risks for researchers after the event.[45]

However, besides these practical risks to the freedom to carry out research, it is important to consider also if the present framework provides appropriate incentives and rewards for researchers who expend significant time and effort in ICT-driven in silico research. This engages the legal area of intellectual property law, and whether the existing rules for rewarding intellectual effort are still capable of protecting researcher investment. Here, one issue concerns the conditions for the copyrightability and ownership of hypermodels; a second issue relates to the initial curation of data that is required in the first place before it is used to build and train the models. From a technical standpoint, data integration is one issue that is still facing the in silico community.

For instance, in cancer research, the study of how individual cancer components interact with each other has led to an explosion in the number of different types of data generated from the patients such as: molecular data, epigenetic data, clinical data, imaging data, pathology data and other laboratory data.[46] These different data types are assembled in order to systematically explore and formalize them in mathematical models. Data integration is key here, but the format, scope, parameter, structure, context, terminology, completeness, etc., of the individual and heterogeneous data are not standardized, which may affect their quality, and ultimately their interoperability and integration.[47] This could also potentially affect collaboration of the different researchers in this field if they use different semantics and techniques to describe, format, submit, and exchange data.

These IP issues that surround the design and development of in silico models, and the question of how far current legal regulation offers adequate protection and rewards for the efforts of the researchers and modelers, will be assessed further in Chapter 6 below.

[43] See Deliverable D4.3.1, for the details of how the CHIC data protection framework operates.

[44] See W. Lowrance, Learning from Experience: Privacy and the Secondary Use of Data in Health Research.

[45] However, as discussed in Chapter 4, the pending GDPR reforms may offer opportunities to establish more flexible and consistent practices moving forward.

[46] Coveney et al (2014), at p. 149 ff.

[47] Ibid.

4 Regulating the Use of Patient Data for Research

4.1 Introduction

As discussed above, achieving more efficient health care, by targeting treatments to the patients best able to benefit from the specific therapy, will require ICT decision-support based on as much relevant data as possible.[48] Thus, in the context of in silico decision support models as in the CHIC project, the model development and validation process falls into several discrete stages, in which researchers, modellers, and clinicians will all require to access and use health data; initially, the models are developed using retrospective data of patients. The data may for example cover patients previously treated for nephroblastoma; the task of the modellers is to develop and test algorithms that correlate differential attributes within that patient population to the target value of interest (tumour shrinkage following chemotherapy). Later there will be further steps for validation, adaptation, and optimization of models, in which data collected prospectively from new patients is required for comparing the real response of the patient to the treatment with the response predicted by the model.

The present Chapter will investigate the legal issues surrounding this kind of use of patient data for medical research and analyse where improvements, with a view to balancing the individual's privacy against the needs of research, might be possible. As noted in Chapter 3, medical confidentiality is a core aspect of medical ethics, forbidding the physician who becomes party to private information when treating the patient from disclosing this to others, apart from in exceptional, justified circumstances. As also noted, such an obligation also rests upon researchers pursuant to the Declaration of Helsinki. In the last 40 years, these traditional duties have been augmented by general duties in respect of fair and lawful information processing, as set in legal rules of data protection.

Like the older rules of confidentiality, data protection rules may be seen as embodying and safeguarding core ethical principles of autonomy, dignity and privacy; they are about making sure that persons remain able to decide how their data will be used and are not exploited or instrumentalised through opaque data processing practices; in addition, it is essential, given the sensitivity of medical data and the risk of harm and distress from unauthorised disclosure, to ensure that data storage and processing is subject to full technical and organisational security. It is apparent that these matters are essential in order for patients to have trust in innovative eHealth applications.[49]

[48] COM eHealth Task Force (2012), 'Redesigning health in Europe for 2020', Section 1.

[49] E.g. in its Green Paper on mHealth, at p 8, the Commission cited evidence that 45 per cent of consumers are concerned about unwanted data processing in relation to the use of mobile health apps.

4.2 The present rules under the Data Protection Directive

At present, the centerpiece of the European regulatory data protection framework is still the Data Protection Directive 95/46/EC,[50] which concretises the fundamental right to data protection set out in Article 8 of the EU Charter.[51] Its scope is "*the processing of personal data wholly or partly by automatic means, and to the processing otherwise than by automatic means of personal data which form part of a filing system or are intended to form part of a filing system*".[52] In this context, personal data is defined as:

> "*[A]ny information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity*."[53]

The Directive requires Member States to ensure – by imposing obligations on entities ('data controllers') who process personal data - that the processing satisfies tests of lawfulness and fairness. In the case of sensitive data, such as health data, Article 8 of the DPD in fact sets out a prima facie prohibition on processing, which is then subject to limited derogations (allowing processing), notably where the individual subject has explicitly consented to this (Article 8(2)(a)); where the processing is required for the purposes of preventive medicine, medical diagnosis, care or treatment of the subject, and is carried out by a (health) professional subject to the duty of confidentiality (Article 8(3)); or where relevant Member State law permits processing for reasons of substantial public interest while providing suitable safeguards for the data subject (Article 8(4)).

In addition, Article 6 DPD, enunciates the relevant principles of 'fair processing'. These include the need for data to be collected for specified, explicit and legitimate purposes and not processed in a way incompatible with those purposes; the data must further be adequate, relevant and not excessive, and not be kept in a personally identifiable form for any longer than necessary. Moreover, under article 17 the processing must be subject to appropriate technical and organizational measures to protect the data against accidental or unlawful access, destruction or loss; that Article goes on to define the level of security required in such cases by reference to "*the risks represented by the processing and the nature of the data to be protected*".

In fact, the DPD is set to be replaced in the next two years by the General Data Protection Regulation (GDPR).[54] This instrument, being a Regulation (directly applicable in Member State law) as opposed to a Directive (which needs to be transposed by Member States in their own law) also offers greater potential in terms of establishing consistent rules. This is significant for health research, as one problem currently faced by multi-site research consortia spread across the Union is that of multi-jurisdictional data protection compliance

[50] See the Directive text: [eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML]. The right to data protection is itself a key aspect of the general right to privacy embodied in Article 8 ECHR.

[51] The background and provisions of the Directive are discussed in detail in Deliverable D4.1, Chapter 4.

[52] Article 3 (1) 95/45/EC.

[53] Article 2 (a) 95/45/EC.

[54] See for the leaked version upon which this analysis rests: [http://www.statewatch.org/news/2015/dec/eu-council-dp-reg-draft-final-compromise-15039-15.pdf].

due to differing implementations of the DPD. As just noted, under Article 8(4), the safeguards accompanying research use of sensitive data are a matter for Member States, and each has enacted different requirements in this regard. For example, when it comes to the requirements for consent to such data use, some Member States require written consent, others do not; and some require prior approval of the processing operation by the national supervisory authority, while others do not.[55]

As the DPD was already extensively analysed in Deliverable D4.1, the rest of this Chapter will focus on analyzing the relevant provisions of the new GDPR and their impact on the use of patient data for medical research and not on the DPD, except as the new provisions will change the framework established by the earlier instrument.

4.3 The new General Data Protection Regulation

The GDPR was originally proposed in a draft issued by the European Commission ('COM') in early 2012. Subsequently, the European Parliament (Parliament) approved its own amended version in 2014 and the Council adopted its version June 2015. Following this, the three EU institutions entered into the stage of the legislative process known as the "Trilogue", involving negotiation between their respective representatives to reach consensus on the text of the GDPR. This text was informally agreed on by the three institutions in December 2015, and is currently being checked and finalized by the EU's legal services. An official vote on its adoption expected in Spring 2016, with the Regulation's likely entry into force set for Spring 2018. In the following we take the latest, unofficial (December 2015) text as the basis for analysis, albeit this may still be subject to final changes. One (non-substantive) change will be with respect to the numbering of the provisions, as some recitals and articles were deleted, and others added, during the extensive legislative process.

The basic principles of the GDPR are very similar to those in the DPD. In this regard it requires that processing adheres to the principles of: lawfulness, fairness and transparency; purpose limitation; data minimization; accuracy; storage limitation; and integrity and confidentiality.[56] The first of these means that any data processing must be based on a legal basis (lawfulness), that the processing must take the interests of the data subject into consideration (fairness), and that the processing must indicate the data to be processed, the reason therefore, and the parties involved (transparency).

The purpose limitation principle places a restriction on the further use of personal data beyond the purpose for which it was collected. This should prevent personal data from being used in perpetuity by the data controller for reasons that the data subject might never have considered. Without it, data legitimately collected, for example from a customer of an online supermarket, could be processed again after the transaction was concluded, for example by a medical insurance company to evaluate the eating habits of the data subject in order to calculate insurance premiums.

Data minimization is closely tied to the purpose limitation principle. By collecting only the minimal amount of data necessary for the purpose, data minimization leads to a *de facto*

[55] See further the discussion in Deliverable D4.1.

[56] Article 5 (1), GDPR.

implementation of the purpose limitation principle – at least insofar as some purposes cannot be achieved without having more data at hand. The same idea can be applied to the principle of storage limitation. By storing data only as long as it is needed, a data controller can automatically prevent any future, further processing of the data after the original purpose was achieved.

Accuracy is another principle that must be seen in relation to the other principles. It is tightly connected with the fairness principle. Personal data is processed for a reason – the outcome of the processing typically affects the data subject in one way or the other. Only where accurate data is processed can the fairness principle be adequately honored. Further, the principle of integrity and confidentiality places upon the data controller, by virtue of her role of custodian of the data, the obligation to take measures to protect the data against unauthorised access and accidental damage and loss through technical and organisational measures. The GDPR also introduces the concept of "accountability" to Union data protection law, stating that the controller is responsible for compliance with the principles and must be able to demonstrate such compliance, too.[57]

In addition to these basic principles that apply to all types of personal data, the GDPR, like the DPD before it, sets further restrictions when it comes to the processing of sensitive data, including health data.[58] Principally, the processing of such data is prohibited.[59] Only under certain circumstances, such as explicit consent by the data subject for specified purposes, may such data be processed.[60] However, Member States are permitted to derogate from the general rules in specific cases.[61]

Other important concepts introduced by the GDPR are the definitions of pseudonymisation,[62] which is defined as "the processing of personal data in such a way that the data can no longer be attributed to a specific data subject without the use of additional information, as long as such additional information is kept separately and subject to technical and organisational measures to ensure non-attribution to an identified or identifiable person"; of "genetic data", "biometric data", and of "data concerning health"[63]. These concepts were undefined/ unmentioned in the DPD, although generally understood to fall under the category of personal data.[64]

In addition, the GDPR introduces further ex ante safeguards for higher risk processiong operations, through the linked mechanisms of a data protection impact assessment ('DPIA') and prior consultation. According to Article 33, a DPIA should take place for:

[57] Article 5 (2) GDPR.

[58] Article 9 GDPR.

[59] Article 9 (1) GDPR.

[60] Article 9 (2) (a) GDPR.

[61] Article 9 (2) (a), (5) GDPR.

[62] Article 4 (3b) GDPR.

[63] In Articles 4 (10)-(12) GDPR.

[64] Article 29 Data Protection Working Party, Opinion 03/2012 on developments in biometric technologies, at [http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2012/wp193_en.pdf].

> "*a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, [that] is likely to result in a high risk for the rights and freedoms of individuals.*"

A further indicator mentioned by the Article for such an assessment, is where large amounts of sensitive personal data are to be processed. Article 34 then states that, if the DPIA shows that the processing would pose high risks to the data subject, absent mitigating steps by the controller, the latter must consult with its relevant national data protection supervisory authority before commencing the processing.

With regard to processing of data specifically for research, the GDPR has a number of relevant provisions that modify the above mentioned principles to favor scientific research subject to certain data processing safeguards. Thus the purpose limitation principle is relaxed insofar that Article 5 (1) (b) GDPR permits further processing of personal data for scientific research purposes. Similarly, Article 5 (1) (e) GDPR makes an exemption to the storage limitation principle for scientific research purposes.

Here, as regards the use of sensitive data for such purposes, Article 9 (2) (i) GDPR provides that this is permissible, in accordance with Article 83 GDPR, where the processing is proportionate to the aim pursued and where "the essence of the right to data protection" is respected and where "suitable and specific measures are taken to safeguard the fundamental rights and the interests of the data subject". In this regard, Article 83 (1) GDPR specifies that:

> "*Processing of personal data for archiving purposes in the public interest, or scientific and historical research purposes or statistical purposes, shall be subject to in accordance with this Regulation appropriate safeguards for the rights and freedoms of the data subject. These safeguards shall ensure that technical and organisational measures are in place in particular in order to ensure the respect of the principle of data minimisation. These measures may include pseudonymisation, as long as these purposes can be fulfilled in this manner. Whenever these purposes can be fulfilled by further processing of data which does not permit or not any longer permit the identification of data subjects these purposes shall be fulfilled in this manner.*"

In such cases, Art 83(2) goes on to provides for the possibility of derogations from the usual rights of data subjects, set out elsewhere in the Regulation (under Articles 15, 16, 17a and 19), to access data concerning themselves, rectify it, or restrict or object to its use, insofar as the exercise of such rights would seriously impair the research objective. These derogations may be enacted either at Union or Member State level.

At the same time, a general competence is preserved to Member States to enact specific rules governing health data processing by the terms of Article 9 (5) of the Regulation. According to this:

> "*Member States may maintain or introduce further conditions, including limitations, with regard to the processing of genetic data, biometric data or health data*."

This may lead to a situation where the processing of such data in different Member States remains subject to different national laws, continuing the fragmentation that exists under the current Directive. Although this risk is to some extent offset by Recital 42a, stating that Member State laws regulating these types of data should not hinder the free-flow of data in

the EU (see below), it is difficult to predict how the balance may be drawn, or the implications for transnational research projects.

Other recitals to the GDPR shed further light on the overall policy approach to using data for scientific research. Recital 25aa notes that because it is not always possible to fully identify the purpose of a processing when it comes to scientific research already when the data is being collected, data subjects should be allowed to give a (broader) consent that permits the use of data for specific areas of research. This is an important concession towards researchers – under the current regime of the Data Protection Directive it is unclear whether such a broad consent is possible. Without broad consent, secondary research becomes generally impossible because valid informed consent cannot be given. However, this will not help any retrospective research with already existing datasets.

As previously noted in Chapter 3, where the exact purpose of the intended data processing is as yet only generally known, eg "cancer research", the data subject is unable to provide specific informed consent to the processing. That would only be possible where the research purpose is further specified. Some jurisdictions, such as the UK, are more receptive to an interpretation that permits broad consent, whereas others, such as Germany, are not.[65] Recital 40 gives a further nod to researchers by clearly stating that further processing for scientific purposes should be considered lawful processing operations. Support of processing for scientific research purposes is also given in Recital 42, which encourages Union or Member State derogations from the general prohibition on the processing of special categories of data, and therefore also personal health data, for scientific research purposes. Recital 42a refines this statement with regard to the processing for health-related purposes. Such processing for scientific research purposes must meet an objective of public interest. At the same time, it also permits Member States to introduce further conditions, including limitations, for the processing of genetic data and health data, albeit only as long as they do not hinder the flow of such data within the Union.

Picking up on the idea of data flows, Recital 88 addresses the transfer of data outside of the EU based on compelling legitimate interests of the controller and mentions that with regard to transfers for scientific research purposes that are not repetitive and that only concern a limited number of data subjects, the legitimate expectations of society for an increase of knowledge should be taken into consideration. Recital 125 reiterates the principle of data minimization when processing personal data, and with regard to the further processing for scientific research purposes states that, where possible, identifiability of the data subject should be prevented through technical measures such as pseudonymisation. Moreover, Recital 125 explicitly states that processing for scientific purposes should comply with other relevant legislation, such as on clinical trials, while Recital 126b notes that the relevant provisions of the Clinical Trials Regulation should apply for the purpose of consenting to the participation in scientific research activities in clinical trials. Recital 125aa reiterates the decision that personal data can be processed for scientific purposes subject to appropriate conditions and safeguards laid down in Union or Member State law, and mentions the benefits that be derived from using personal data stored in registries.

[65] For the diverse ways in which the DPD has been interpreted in Member States' data protection laws, see D. Beyleveld et al (eds), 'The Data Protection Directive and Medical Research Across Europe'.

Other definitions found in the GDPR generally are the same or very similar to those of the DPD. For example, the concept of "personal data" is still understood as "any information relating to an identified or identifiable natural person".[66] Going beyond the Data Protection Directive, however, the GDPR gives examples of identification, "such as in particular by reference to an identifier such as a name, an identification number, location data, online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that person".[67] Recital 24 GDPR clarifies that online identifiers such as IP addresses, cookie identifiers, etc. *may* identify an individual and may therefore be personal data. The Parliament's proposal, in contrast, took the approach that such identifiers always are personal data, unless they do not lead to identifiability of an individual.

A facilitation of certain types of research is made by Recital 23aa, which explicitly excludes the data of deceased persons from the scope of the GDPR, albeit with derogations possible for Member States. The DPD was silent on the matter. Some Member States addressed the issue, either including or excluding the data of deceased persons from the scope of their national data protection legislation implementing the DPD, while others were silent on the issue. While it is encouraging that the GDPR addresses the issue, the possibilities of Member State derogations means that harmonization across the Union will not be achieved. The general exclusion of the data of deceased persons, however, would be beneficial for medical research and in particular *in silico* research with its concentration on retrospective data – data which will quite often also refer to deceased individuals.

[66] Article 4 (1) GDPR.

[67] Article 4 (1) GDPR. See also Recital 24 GDPR.

4.4 Implications for E-health, including in silico research

In the CHIC position paper prepared by LUH following the COM draft proposal Regulation, and circulated within the VPH community in summer 2013,[68] three main issues were noted that have caused researchers using health data difficulties under the framework of the DPD:

> "*Terminological and factual difficulties in deciding if some given data qualify as personal or not, so that the legal DP framework applies notionally or not (a related ambiguity concerns how far processing of such data by anonymising or key-coding them are themselves operations that trigger the full legal obligations under that framework);*
>
> *Independent of the above, difficulty in deciding when the patient data subject should be approached for consent to data use, and if so what would count as appropriate consent in a given case;*
>
> *Independent of the above, difficulty in deciding when it would be appropriate or even mandatory to approach a review body of some kind, e.g. an ethics committee or data protection authority, for authorization of the proposed research using the data.*"

In this section we evaluate how far the GDPR in the current form, likely to be enacted, addresses these problems. First, one issue that did not change compared to the DPD is that the act of anonymising personal data is a type of processing. While the processing of anonymous data itself, i.e. "information which does not relate to an identified or identifiable natural person or to data rendered anonymous in such a way that the data subject is not or no longer identifiable", is outside the scope of data protection law,[69] that is not the case for the act of transforming personal data to "data rendered anonymous".

For research purposes that could be conducted with anonymous data, but where there was a lack of anonymous data sources, this presented a serious hindrance. However, because Articles 5 (1) (b), 83 (1) GDPR permit the further processing for scientific purposes, this is arguably no longer such an issue and should facilitate certain types of research. While the DPD generally permitted the same, subject to suitable safeguards,[70] the leeway granted to Member States with regards to the safeguards meant that the further processing of personal data for scientific purposes was easer in some jurisdictions than in others.

More generally, as noted, Article 83 GDPR emphasises the importance of the data minimization principle when data is used in research. Ideally anonymous data should be used; however, where this is not feasible for achieving the research purpose, personal de-identified data may be used. Broadly this mandates technical (such as pseudonymisation, where appropriate for the purpose) and organisational measures to give practical effect to the principle of data minimization. At the same time, as the Regulation clarifies in Recital 23a, pseudonymisation is only a tool to reduce the risks for data subjects and that it helps "*controllers and processors meet their data protection obligations*"; accordingly other data protection measures are not to be seen as precluded, but should be used in addition to maintain data security.

[68] Appended to CHIC Deliverable D4.1, as Annex 2.

[69] Recital 23 GDPR.

[70] Articles 7 (e), 8 (4) DPD.

Pseudonymous data can be protected by applying legal, organizational and technical safeguards that prevent the data subject from being re-identified, thereby leading to what may also be termed 'secure reversible de-identification'. This also has the benefit that, exceptionally, it might be ethically appropriate to link back, e.g. where data-mining uncovers information of vital importance for the well-being of a particular patient. The further legal safeguards may include contractual prohibitions on researchers in a given project from using the data for other than the strict research project purpose, on disclosing the data outside the researcher group or seeking to re-link the data to patient subjects either by the key-code (except in exceptional circumstances, as mentioned) or otherwise.[71]

The organizational and technical safeguards should include strict data access policies and controls on individual data users, state of the art secure servers, and encryption of data during transit so as to make access by unauthorized persons virtually impossible. A specific issue here (and also later when deploying hypermodels) is that, where the volume of data processing requires the use of a cloud-based infrastructure, it will be necessary to have safeguards in place to ensure authorized users retain exclusive control of the data; in normal circumstances this means a private rather than public cloud solution should be adopted.[72]

Secondly, as regards consent issues, as we saw the permissibility of obtaining broad consent is anticipated in Recital 25aa for cases where it is not possible at the point of collection to fully specify further scientific purposes for which the data may be suitable. This may be regarded as an advance in terms of freeing up data for additional research uses over the Directive, which stipulated the need for specific consent. Moreover, in cases of retrospective data previously collected and held by the researcher, it appears that the need for reconsent may be dispensed with more readily than under the Directive; then, where it was logistically impractical or otherwise undesirable to recontact subjects or patients, the researcher was required to justify the processing on the alternative basis of substantial public interest under Article 8(4) of the Directive.

By contrast, it appears from the text of the Regulation that, subject to the adoption of proper safeguards in line with Article 83, a researcher may reuse data for further research (different from the original research purpose) so long as the new research is 'not incompatible' with the original research. Indeed, in principle this also extends to the use of data for research, which was not originally collected for research (but a different purpose, such as diagnosis or treatment). The critical provision here, is Article 6 (3a) of the Regulation, which also provides guidance as to relevant factors in determining when a proposed new use is not incompatible with the collection purpose. These factors include the context in which the data were collected, the possible consequences of the further processing for the data subject, and the existence of suitable safeguards, which may include encryption or pseudonymisation.

Thirdly, as regards approval issues, it was noted in the 2013 CHIC position paper[73] that researchers on the ground often wish, for a variety of professional, ethical, and prudential reasons, to obtain some form of review and approval to their use of health data for research

[71] Such a contractual framework has been deployed within the CHIC project: see further Deliverable D4.3.1 (May 2014).

[72] See the 20012 Opinion of the Art 29 WP on Cloud Computing, WP 197.

[73] See note 64.

(even where the data is not obviously linkable to individual subjects, and/or where the subjects have agreed to the use). Indeed, as discussed in Chapter 3 of this Deliverable, the need for such review – certainly if there are doubts as to subject consent - is set out in key international ethical documents, e.g. the Declaration of Helsinki (Principle 25).

Admittedly, under the DPD the possibility for Member States to require prior checking/approval of particular processing operations by national supervisory authorities has been allowed for under Article 28; and some Member States have made use of this as a safeguard under article 8(4) before permitting sensitive data processing in the public interest. However, other States did not; instead, the general desirability for independent oversight, typically by LECs, has been promulgated in diverse soft law guidance. Typically, though, this occurs at local level, which has led to problems of delay and inconsistency, where multiple oversight bodies for different would-be data providers must each be approached in relation to a single piece of informational research.

Here, it appears that the Regulation will improve matters. Thus, under Article 35(1)(c), projects that utlilise large amounts of sensitive data will be required to install a data protection officer to monitor data protection compliance. Moreover, it seems very likely the processing of health data for research will qualify as a processing operation of 'specific risk' under Article 33, triggering the need for a DPIA, and in most cases prior checking by a supervisory authority under Article 34. In this way, the Regulation would serve to institutionalize the independent oversight and prior approval of the proposed data processing by a single competent supervisory body, applying consistent and well-considered criteria.[74]

Related to this, one mechanism that could become a powerful tool to facilitate research within the Union are the codes of conduct that the GDPR encourages (also as a pointer to proper compliance with a relevant DPIA). In this regard Article 38 GDPR foresees "*codes of conduct intended to contribute to the proper application of the Regulation*", the adherence to which shall be monitored by the competent supervisory authority and by bodies accredited by the supervisory authority.[75] Codes of conduct for organisations conducting processing across several Member States (such as medical research) would need to submit the Code to the EDPB. This process can serve as a harmonizing mechanism across the Union by offering a clear guide to data controllers and processors on how to implement the research provisions of the GDPR and because codes of conduct that relate to processing activities in several Member States must be positively assessed on Union level before being adopted.[76]

Overall, then, the GDPR can be said to be favorable towards data processing for scientific research purposes, while respecting the data subject's rights, too. The major 'fly in the ointment' that exists is the possibility for Member States under Article 9(5) to derogate from the Regulation's provisions, thereby undermining the Regulation's potential to harmonize legislation across the Union. On the plus side, the Regulation's encouragement of codes of conduct appears to have significant potential for promoting consistent, good data processing

[74] In the first instance this would be the task of the national data protection authority in each Member State (in the UK's case, the ICO). However, it may be assumed that national authorities would work together in developing consistent guidance and procedures, and where appropriate consult with the EPDB.

[75] Article 38a (1) GDPR.

[76] Article 38a (2b), (3), (4) GDPR.

practice in particular sectors of processing activity. As suggested in Chapter 7 below, this could offer an opportunity to evolve consistent, transnational rules of practice in the health data research sector, including for the VPH and in silico modelling communities. In the light of this, and in the context of its work in the CHIC and related projects, LUH plans to promote further discussion in the VPH community towards developing a tailored code of conduct to cover in silico research-related data processing, and to report on progress made as an aspect of the updated data protection deliverable D4.3.2.

5 Validating In Silico Research, and Ensuring the Safety of Patients Treated via E-Health Decision Support Systems

5.1 Introduction

As noted above, for validation, adaptation, and optimization of models, the real response of the patient to the treatment will need to be compared with the predicted response. Thus once a hypermodel has been built with the potential to guide a specific clinical decision (should the physician give treatment A or B to a given cancer patient with attributes p, q, r, s…?), the question of validating its accuracy arises. Some initial testing will certainly be possible virtually, by running the model on retrospective data available from other patient populations (not used to build the model). However, there will then be the need for legal compliance validation – in terms of ensuring that the model is fit for purpose for treating real prospective patients. In the context of CHIC, once the models have achieved a certain level of maturity/accuracy, they will be subject to preliminary-testing by clinicians.

At this point the patient's individual data at diagnosis is fed into the model (previously trained on retrospective data), and the model will seek to predict that patient's likely response (tumour shrinkage or not) by comparing that particular patient's data values against the data of retrospective patients where the outcome is known. Besides the data protection issues, discussed in Chapter 4, this raises further ethical and legal regulatory issues that are considered in the present Chapter. In this regard we shall consider especially the relevance and potential application of the medical devices regime, which, as in the case of data protection law considered above, is presently the subject of reform initiatives at EU level.

Later, in section 5.4, we shall also seek to anticipate future effects the deployment of in silico models in clinical practice may have on the direct treatment and care of patients, and how far this may give rise to implications and risks that stand in need of further regulation. Though strictly a matter of E-health practice, rather than research, it is clearly important to think already about an appropriate framework for the products of in silico research to be utilized to their full potential, while minimizing associated patient risks. The main issues here concern the possible impact of potentially extremely accurate predictive model on the doctor-patient relationship, including the need for management of information disclosure to the patient and relatives, as well as potential liability risks that may arise. We also assess the adequacy of current liability schemes to address cases where predictions turn out to be wrong, requiring the ascertainment of responsibility under diverse compensation regimes.

5.2 Validating and testing the safety and efficacy of clinical support tools

A general problem faced by E-health diagnostic approaches, including by the VHP community when developing in silico models, is how to translate the research results into daily clinical practice. This centres on what appropriate methodology is necessary for validation of the simulations, and their use as decision support systems. Are these in silico models and platforms medical devices, subject to the medical devices regime? If so, what is the appropriate risk class (determining the type of the testing procedure)? Might there be the need for full clinical trials, similar to those for pharmaceutical products (as discussed in Chapter 3) in order to validate and certify models? Clearly, the answers to these questions will affect the length of time it will take to translate in silico decision support models from laboratory to bedside.

Here the first point to make is that an *in silico* model will very likely qualify as a medical device within the definition of the EU Medical Devices Directive 93/42/EC ('MDD'). The MDD is the most general of three complementary EU Directives enacted in the 1990s that regulate the testing, certification and post-marketing surveillance requirements in Europe for various types of medical device for human use. The other Directives are the Active Implantable Medical Device (AIMD) Directive 90/385/EEC, and the In Vitro Diagnostic Device Directive (IVD) 98/79/EC, which contain the rules for powered devices implanted into the human body, and devices which analyse bodily specimens, respectively. By contrast, as discussed in Chapter 3, the regulation of medicinal products (pharmaceuticals) occurs under the separate regime established by the Clinical Trials Directive 2001/20/EC, whose provisions are due to be replaced by those of the Clinical Trials Regulation EU No 536/2014. The rules for medical devices are also currently undergoing reform, with plans for the three 1990s Directives also to be replaced by two new EU Regulations in the foreseeable future.[77]

Given the focus of the CHIC project, the ensuing discussion will centre on the provisions of the MDD, including their application to medical devices that operate using computer software. First, as to the meaning of 'medical device', this is defined in Article 1 (2) (a) as:

> *"any instrument, apparatus, appliance, software, material or other article, whether used alone or in combination, including the software necessary for its proper application, intended by the manufacturer to be used for human beings for the purpose of: diagnosis, prevention, monitoring, treatment or alleviation of disease, ... and which does not achieve its principal intended action in or on the human body by pharmacological, immunological or metabolic means, but which may be assisted in its function by such means;"*

This is a wide definition that, as can be seen, includes express references to software;[78] moreover COM guidance, issued in 2012, makes it clear that software operating in its own

[77] See the EU Council Progress Report 15881/14, on the draft Medical Devices Regulation, at http://register.consilium.europa.eu/doc/srv?l=EN&f=ST%2015881%202014%20INIT. A common position was agreed in September 2015 but there currently remainsl significant debate on parts of the reforms.

[78] Introduced by Directive 2007/47/EC, amending the MDD.

right may fall under the rules as so-called 'stand-alone software'.[79] Moreover, it was expressly suggested that clinical decision support software will be covered, defined as:

> "*computer based tools which combine medical knowledge databases and algorithms with patient specific data. They are intended to provide healthcare professionals and/or users with recommendations for diagnosis, prognosis, monitoring and treatment of individual patients.*"[80]

As a medical device the hypermodel would be subject to a set of pre-marketing testing and certification requirements. Only devices that go through the relevant testing procedures, and receive a CE mark attesting to this, may be lawfully placed on the European market. In this regard, the MDD (in its Annex IX) undertakes a broad division of medical devices into four different risk-classes, based on their general characteristics. In Article 11, differential testing procedures are then specified for use in respect to each class, whose rigour increases according to the potential for harm to users in case of malfunction.

In the case of in silico cancer models, designed to inform treatment decisions for a serious disease, and where the treatment itself may involve high risks, it seems likely the device would be classified as at least Class IIb: medium high risk. In this case, the testing and evaluation would need to be performed by an independent entity, known as a 'notified body'. These are for-profit companies, licensed in each Member State by the competent national medical devices authorities (set up by relevant domestic law transposing the MDD).

According to the MDD, Annex X, the evaluation that the device satisfies the essential requirements should be based upon 'clinical data'. The aim is to show that the device does not pose undue safety risks to users, and also that it performs in the way the manufacturer claims. At present there is no need to go further and show a positive clinical benefit, in terms of efficacy, from using the device. Related to this (and in contrast to the more homogenous rules for pharmaceutical trials), diverse testing approaches and procedures may be deployed in a clinical investigation (under Article 15 and Annex VIII) to obtain the relevant data. This reflects the diverse nature of products that may qualify as medical devices, which may make a single approach (e.g. use of randomized control trials) less appropriate or practical. However, this may change as a result of the ongoing legislative reform process, aimed at the replacing the MDD by a Regulation; thus the draft proposal[81] would favour the use of more rigorous 'clinical investigations' of devices (similar to the sponsored clinical trials required in the field of medicinal products) in order also to show clinical benefits from use.

[79] See EC, Medical Devices Guidance Document, MEDDEV 2.1/6. 2012, January, available at: [ec.europa.eu/health/medical-devices/files/meddev/2_1_6_ol_en.pdf].

[80] Ibid, p. 20.

[81] See the draft MDR: Council Common Position 2012/0266 (COD) (September 2015), available at: [http://data.consilium.europa.eu/doc/document/ST-12040-2015-REV-1/en/pdf].

5.3 Implications for E-health, including in silico research

As discussed in Chapter 4, E-health clinical support devices, such as in silico models, will be developed and built using retrospective patient data. However, once they reach a level of maturity where the predictions they make (still using retrospective 'hold-out' data) attain an acceptable accuracy, a further stage is for them to be tried in a real clinical environment, where they are fed with the prospective data of newly diagnosed individual patients. At this stage, besides the need to assess their usability by the treating clinician, it will be important to assess how far the predictions with the new patient data compare with the models' previous accuracy levels. At the same time, it would clearly be premature (and potentially negligent) for the clinician to rely already on the models' predictions in determining the treatment plan for a given patient. Instead the idea is that the clinician would simply observe how the model performs, while continuing to base treatment on independent clinical judgement. Only later, after the models have demonstrated their ability to provide accurate predictions for individual cases, would it be proposed to move towards formal validation, with a view to the models being certified as suitable for use in clinical practice.

As discussed above, the relevant regulatory scheme for achieving validation and certification is that for medical devices, set out in the MDD 93/42/EC. However, two specific issues that arise in relation to applying this regime to in silico model testing as just described are, first: when would the existing validation rules be triggered, and what concrete tasks do they impose on the model developers and testing clinicians? The second issue also concerns the potential effect of the impending reforms being enacted, in the form of the proposed Medical Devices Regulation to replace the Directive. As noted, one probable change will be for medical devices to be subject to more rigorous and homogenous testing regime prior to certification, similar to that for pharmaceutical products, and it may be asked if this would be an appropriate development in the case of in silico predictive models.

Regarding the first issue, a question that remains somewhat unclear under the present rules is at what point the observational activities of clinicians, who compare the model predictions with what actually occurs following the treatment they choose for their patients, should be seen as a 'clinical investigation' for the purposes of Article 15 MDD. This provision allows the models to be deployed in clinical practice without yet having the CE certification mark; however, its application at this point may imply a degree of formality and expense (involving the need for notification to competent national authorities and ethics approval) that would arguably be premature in the context of small-scale project-level research. Thus it could mean the observing clinicians are subject to quite onerous clinical data gathering duties,[82] including the need for clinics to have mechanisms to protect 'observer neutrality', such as by entrusting observations of the real patient to separate clinicians to those who are aware of the hypermodel's prediction. Arguably, a preferable interpretation may be to see the clinical trial rules as triggered once the models are made available outside the project for observational testing by external clinicians. The (acceptable) corollary is also that before this point the observational data gathered by the clinicians inside the project would lack probative force as clinical data for MDD purposes.

[82] In line with the clinical evaluation requirements of MDD, Annex X.

This leads on to the second issue, which as noted, concerns what form the further gathering of clinical data to validate the cancer models as medical devices should take, and also how this may stand to change as a result of present legislative reform (leading to a new Medical Devices Regulation). Here we may begin with the point noted in 5.2 that the aim of the validation process is to assure the public that medical devices are safe in practice and perform as claimed by the manufacturer. By contrast (unlike the case of pharmaceutical products), it is not required under the present regime to show efficacy, in the sense that the device performs better than existing devices or therapies. Admittedly, in the context of E-health decision-making support tools such as in silico models, safety and efficacy may actually hard to separate out: thus unless the model leads to a demonstrable benefit in clinical care, by enabling clinicians to make better treatment decisions, it would seem safer (and in better accord with clinical ethics of accountability) to leave the decision to the human clinician alone.

To this extent it may appear the changes mooted in the draft Medical Devices Regulation (MDR), that would require medical devices to demonstrate clinical benefit/efficacy, are quite germane in relation to E-health decision support tools; nonetheless a problem at the same time concerns the more formal methodology generally favoured by the draft Regulation for showing clinical benefit, encouraging the relevant clinical data to be acquired via sponsored studies including RCTs.[83] The rationale and architecture of RCTs has been described in Section 3.2 above. However, as noted, they may be difficult to carry out in the diverse field of medical devices, where inter alia the associated mechanism of 'blinding' patients and clinicians to the treatment process is often hard to achieve, both in practical and ethical terms. This also presents challenges in the context of in silico model validation.[84]

In the latter context, it may also be open to question how far the methodology of an RCT is necessary for gathering informative clinical data (while safeguarding patient/subject safety). In particular such models offer a 'second-level' (virtual) means of choosing the optimal 'first-level' (real) therapy for a given patient, rather than a therapy in their own right. In principle, therefore, it ought to be possible to test them purely by observation – if it turned out, say, after having been run on the data of thousands of nephroblastoma patients, a given model nearly always succeeds in picking out the minority of patients whose tumours will not respond to chemotherapy (where immediate surgery is indicated), then there would be a good case for starting to base decisions on the model. It may be doubted whether, as the intermediate step of running an RCT would be in point, or even ethical here.[85] This is not to under-estimate the considerable, systematic effort that will be needed when validating and seeking to implement hypermodels. As discussed, in the context of formal validation, and to avoid the risk of observer bias, hospitals should ensure that observations of the real patient are not carried out by clinicians aware of the hypermodel's prediction.

[83] See the draft MDR, Council Common Position, Annex XIV.

[84] For further discussion of possible validation methods including trials for in silico models, see Coveney et al, Computation Biomedicine (Oxford University Press, 2014) at 239 ff.

[85] The 'equipoise' in medical knowledge that gives randomisation its ethical basis would be lacking, where the observational data suggests treating the control arm of patients conventionally (chemotherapy first) will harm the interests of a significant minority of them. Other matters such as the usability of the models in clinical practice, would also need testing, but it is unclear again if randomisation would be appropriate there.

Another significant special feature is that *in silico* simulations performed by hypermodels, augmented by image visualization functionality, may often allow the clinician to view the patient's predicted progress in a chronologically contiguous manner, akin to repeat observations of that patient in real-time. In this situation there would be little or no delay in verifying the accuracy of a prediction: the clinician, by observing the patient, can see quickly if the actual course of events following treatment recommended by the hypermodel conforms to, or diverges from, what the model indicated would happen. In the case of a divergence, the clinician would be able to switch away quickly from using the (inaccurate) model and adopt a different treatment. Admittedly, a problem would remain in cases where adopting the treatment indicated by the hypermodel would involve a radical intervention contrary to the standard therapy (e.g. immediate surgery on a nephroblastoma patient). In such a situation, it is suggested the clinician should adopt the radical course only if this tallies with his own clinical experience and judgment, informed also by a good knowledge of the hypermodel's success in predicting outcomes in the relevant patient population.

Subject to the above points, and also always assuming informed consent from the patient or his legal representative (who should be told the clinician is – at least in part – basing the treatment on an automated algorithmic mechanism), hypermodel validation by clinicians could arguably be classed as a case of innovative therapy for the benefit of the individual patient. If that is so, there would also be an argument for promoting/maintaining flexible rules in this area, that derogate (at least in part) from the generic approval regime under EU medical devices legislation. This could occur by introducing specific rules tailored to in silico model testing, or by more generic exemptions, invoking e.g. 'compassionate use' grounds (as found in the US for unapproved pharmaceutical products)[86] or classifying a model when populated with a specific patient's data as a custom made device.[87]

[86] [http://www.cancer.org/treatment/treatmentsandsideeffects/clinicaltrials/compassionate-drug-use].

[87] The latter would, as under the MDD, remain exempt from the MDR standard approval requirements.

5.4 Regulating the use of in silico decision tools in clinical practice

It is apparent that *in silico*-based medicine has the significant potential to improve healthcare delivery; however, it also poses some legal and ethical challenges when applied in treatment scenarios. This will require clarifying important ethical boundaries, including the risk that reliance on *in silico* predictions may estrange or otherwise affect the physician-patient relationship. How much should one rely on leaving potentially vital decisions to an automated system that may not have the ability to appreciate the unique character and personality of every individual that doctors gain from physical interaction, training and years of experience? Admittedly, these issues concern clinical deployment rather than the prior research and validation of models, so are strictly speaking outside the remit of the present report. However, since they will influence the receptivity of the market for E-health innovations such as in silico models, and thereby impact on stakeholder investment risks, we outline key issues here: in same cases these may also invite regulatory response at European level.

In the first place, as with other novel diagnostic interventions, it will again be important to secure the patient's informed consent to be treated with the aid of the hypermodel. As part of this, the patient should, to the extent that the clinician bases the decision which treatment to provide on the model's prediction, be made aware of this. However, it remains unclear how specific the consent would need to be (e.g. in explaining the logic underlying the decision) to be legally valid. The law would presumably have to take account of the practical difficulties clinicians and patients may have in giving and understanding a detailed explanation.

Challenges may arise too for the doctor's ethical duty of candour towards patients. Assuming for example that a model tells the doctor that any course of therapy will be hopeless for a certain patient, how should the doctor act on this information?[88] Such a scenario may also raise difficult distributive justice questions if models were to later include functionality for computing cost-effectiveness of different treatment options, or indeed determine, between patients, who would be the most efficient recipient of some resource-intensive therapy.

A further issue is that of legal liability in the event of adverse outcomes resulting from inaccurate or incorrectly interpreted models or data. Admittedly, the model can only give probabilistic information, but clearly if it gives a wildly wrong prediction the doctor may end up taking a decision he would not have taken otherwise. In this case, who should be liable – the doctor, the model-developers? The private law rules of tortious or contractual negligence will here be relevant, but may fail to vindicate the interests of injured patients due to the requirement for the patient to show fault on the part of one of the multiple actors involved. In this regard, the STEP project consortium has identified potential factors that may lead to an unforeseen adverse outcome in VPH / in silico models, such as: patient

[88] See further, I Cohen, et al, The legal and ethical concerns that arise from using complex predictive analytics in health care.

variability, databases populated with incorrect data, inappropriate use of data, the use of a flawed model, a misunderstanding of the assumptions associated with a model, etc.[89]

As an alternative, the strict liability based product liability regime may here appear a preferable model. This is particularly so as the regime is founded upon an EU Directive,[90] and thus possesses greater uniformity than private law liability rules, which are the product of diverse Member State law. In EE-health mediated transactions, potentially involving actors from several Member States, a consistent approach to offering redress - which, on the other side, also clarifies the liability risks for the health service provider - would be a considerable advantage. However, presently there is uncertaintly and debate (reflected in differential practice across Member States) as to how far the product liability regime applies to computer software programs in the first place.[91]

Lastly, and more broadly, the advent of personalised health care may present some significant challenges (of an ethical and social nature) to traditional conceptions of health insurance and the market for it. The very idea of insurance is to mitigate risks of uncertainty about the future. Many pay to mitigate the risk only few will be affected by, due to the fact that nobody knows in advance who will be lucky and who will not. The more, though, that it becomes possible to accurately predict the future development of a given individual's state of health, the less may insurance, as a form of risk-pooling, seem appropriate. Persons, with a favorable risk profile might prefer to opt out (instead paying directly for their care as the need arises); and conversely, persons with poor predicted health outcomes may find no one willing to insure them. Is this ethically, socially and legally justifiable? It will be important to consider policies that address and mitigate these wider social risks and challenges through a sound regulatory, legal response that finds a justifiable equilibrium between autonomy, public interest and human rights.[92]

[89] STEP Consortium, Seeding the EuroPhysiome: A Roadmap to the Virtual Physiological Human, p.80, available at: [www.ncbi.nlm.nih.gov/pubmed/18559316].

[90] Council Directive 85/374/EEC of 25 July 1985 on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products.

[91] See further, K Alheit, The applicability of the EU Product Liability Directive to software; L. Vihul, The liability of software manufacturers for defective products.

[92] For an interesting recent approach, which seeks to preserve autonomy through a revitalised conception of active informed consent, see AN Sanchez,'Privacy by default' and active 'informed consent' by layers: essential measures to protect ICT users' privacy'.

6 Intellectual Property Rights: Rewarding Researcher Efforts in E-Health Innovation

6.1 Introduction

In this Chapter, we consider how the works of cancer modeling, and in particular amalgamation of models into multiscale hyper-models, may be protected in terms of IP law. This concerns the types of IPR that may apply in principle (insofar as relevant requirements are met) and the chances and benefits for E-health researchers of obtaining the protected rights. The key underlying issue for the purposes of this report is whether the existing rules for rewarding intellectual effort are adequate and appropriate, in this innovative research area, in order to recognize and protect relevant researcher investment. In this context it is useful to begin by considering the various stages that underlie the creation of an E-health related research product, such as an in silico hypermodel.

Such a model is the culmination a lengthy and complex development process, requiring considerable efforts by multiple parties. Thus In the context of the CHIC project, the research work on multiscale cancer modeling, aims at "*rational, coherent and comprehensive exploitation of the invaluable information hidden within human multiscale biological data*"[93.] The work, as with E-health research in general, relies on – and begins with - the collection of large amounts of retrospective health data from past patients. It is this data that provides the raw material for mining and analysis that yields decision-tree and other patterns, such as from clustering, that are then embodied in algorithms that drive the different scientific models.The latter are defined as: "*finalized cognitive constructs of finite complexity that idealize an infinitely complex portion of reality through idealizations that contribute to the achievement of knowledge on that portion of reality that is objective, shareable, reliable and verifiable*."[94] These scientific models correspond to the biological processes, being simulated.

The scientific models are then implemented in silico via computer models. At this stage, the scientific models, selected for simulation, are encoded into computer models. In the context of CHIC, a computer model is defined as: "*a computer program that implements a scientific model, so that when executed according to a given set of control instructions (control inputs) computes certain quantities (data outputs) on the basis of a set of initial quantities (data inputs), and a set of execution logs (control outputs).*"[95] In this process, either the already developed codes of tumor models are broken down into simpler models or computer codes of elementary biological processes (biomechanics) are developed anew. Coding finalizes the development of cancer models.

A further stage then is for the elementary models, each of which represents a biological process at a single scale, to be combined into multiscale hyper-models, so that spatiotemporal simulation of the clinical trials and studies, addressed by the CHIC project, is

[93] CHIC EC-GA, Annex I, DOW, Abstract, p.3.

[94] M. Viceconti, 'A tentative taxonomy for predictive models in relation to their falsifiability'.

[95] CHIC, Deliverable No. 7.1, "Hypermodelling specifications", CHIC, Downloads, Deliverables, Public Deliverables, <http://chic-vph.eu/uploads/media/D7-1_Hypermodelling_specifications.pdf> 2016-02-07.

achieved.[96] In the context of the CHIC project, hyper-models comprise "*choreographies of component models, each one describing a biological process at a characteristic spatiotemporal scale, and of relation models/metamodels defining the relations across scales*."[97] The aim thereby is that in a given instance (e.g. when populated with the data for a given patient) the hyper-model is able to reproduce complex biological processes, involving multiple phenomena which are respectively captured in single models; it "*emerges from the composition and orchestration of multiple hypomodels, each one of which is capable of simulating a specific entity or phenomenon… and can simulate an entity or phenomenon that may be more complex than the ones simulated by each separate simpler model*."[98]

On the basis of the above, there are a number of questions that arise in relation to IPR. In the first place, what protection (if any) might be available to the preliminary work of collecting and curating data? This is an issue of general significance for any type of E-health-related research, and how relevant work here may be encouraged and rewarded. In the second place, there are issues more specific to in silico modelling, including the work of amalgamating models into hypermodels that occurs in CHIC. The question here is how far the researcher and modeler efforts may obtain protection under the traditional schemes of IPR: could a hypermodel developed through in silico research and modeling qualify as an invention subject to patent law protection? If not how far might copyright protection, based on the fact that the models are expressed in software code, provide protection (including for the previous steps that led up to the software development)? Below we will consider these various issues in turn.

[96] CHIC Grant Agreement no: 600841, Annex I - DoW, WP6.

[97] "Computational Horizons In Cancer (CHIC): Developing Meta- and Hyper-Multiscale Models and Repositories for In Silico Oncology", CHIC, Project, Objectives, <http://chic-vph.eu/project/objectives> 2016-02-17.

[98] CHIC, Deliverable No. 7.1, "Hypermodelling specifications", CHIC, Downloads, Deliverables, Public Deliverables, <http://chic-vph.eu/uploads/media/D7-1_Hypermodelling_specifications.pdf> 2016-02-07.

6.2 Rewarding data curation

6.2.1 Data curation in CHIC

As noted, a key issue relates to protecting efforts made in the initial curation of data that is required in the first place before it is used to build and train the models. From a technical standpoint, data integration is still a significant challenge for E-health research, including the in silico community. Thus, in cancer research, the study of how individual cancer components interact with each other has led to an explosion in the number of different types of data generated from the patients such as: molecular data, epigenetic data, clinical data, imaging data, pathology data and other laboratory data.[99] These different data types are assembled in order to systematically explore and formalize them in mathematical models.

Data integration is key here, but the format, scope, parameter, structure, context, terminology, completeness, etc., of the individual and heterogeneous data are not standardized, which may affect their quality, and ultimately their interoperability and integration.[100] This could also potentially affect collaboration of the different researchers in this field if they use different semantics and techniques to describe, format, submit, and exchange data. The curation required here to ensure the data relates to and measures the same phenomena with sufficient accuracy to be usable is a large and painstaking task. It includes the problem of dealing with incomplete data fields and cross-checking that various indices were measured and recorded in a similar way (e.g. images were taken using similar equipment, co-morbidities were classified using the same terminology, etc). It is evident too that considerable expertise and skill is required for it to be performed well: the curator needs to have a real feel and understanding for the subject matter in order to make sensible judgements in resolving various gaps and uncertainties.

Thus in CHIC the development and validation of cancer models is based on the use of clinical information either taken from the literature or provided by the clinical partners, in particular, USAAR, KU Leuven, UNITO.[101] For instance, the clinical trials HGG-IMMUNO-2003, HGG-2006, HGG2010 are supposed to serve as data providers for the development of a 4D glioblastoma treatment response model. A molecularly enhanced 4D multiscale model of non small cell lung cancer (NSCLC) response to chemotherapeutic and/or radiotherapeutic treatment is developed based on the data provided by USAAR. The model of prostate cancer is to be validated against the clinical data of patient database, residing in the Institute for Cancer Research and Treatment (IRCC) in Turin, Italy.[102]

However, whereas the clinical parties USAAR and KU Leuven manage their clinical information and store the results of clinical trials using ObTiMA, an ontology-based clinical trial data management system[103], the data management system, used by UNITO (as is the case for many other medical institutions) is specific to the medical activities conducted by UNITO. The use of diverging data management systems by the clinical institutions lead to

[99] Coveney et al (2014), at p. 149 ff.
[100] Ibid.
[101] CHIC EC-GA, Annex I, DOW, WP6, pp.31-33.
[102] Ibid
[103] Stenzhorn H, et al, 'The ObTiMA system - ontology-based managing of clinical trials'.

the situation that the data, provided by different institutions, is not inter-operable with each other and mostly may not be used for research as such. The clinical data also needs to be post-processed by the modelers so that it fits into the set of parameters, which the models recognize and can utilize for running the simulations. The inputs, outputs and descriptions of processes, simulated by the models, are standardized into the set of parameters, acceptable and usable by all cancer models.[104]

6.2.2 Potential protection of IP in data and information

At the present stage of the European legislation, there are several potential schemes of legal protection that may come into question as options for protecting data collected and used in E-health research. These include the sui generis database right; know how protection, and related rights. However, as we analyze below, these types of protection are tailored to protect specific objects (data repositories, confidential information with commercial value, etc) and, as we discuss, none as such guarantees adequate protection to protect the investment, made in curating the clinical data for research. We consider these schemes, and illustrate their application by reference to the data used in CHIC, in more detail below.

6.2.2.1 The sui generis database right

The legal protection of databases is provided by the Directive 96/9/EC of 11 March 1996 on the legal protection of databases (the Database Directive).[105] Such protection is granted in recognition of the fact that constructing a database requires "*investment of considerable human, technical and financial resources*"[106]. The directive aims to reward and protect such investment by providing the maker of a database with a sui generis data base right that places him in a position to prevent unauthorized access and copying of the database contents, which he compiled. In this regard, Article 7 Database Directive states:

> "*Member States shall provide for a right for the maker of a database which shows that there has been qualitatively and/or quantitatively a substantial investment in either the obtaining, verification or presentation of the contents to prevent extraction and/or re-utilization of the whole or of a substantial part, evaluated qualitatively and/or quantitatively, of the contents of that database*." The object of protection in terms of the Database Directive is a 'database' meaning "*a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means*."[107]

Protection of databases by the sui generis right can be considered as a plausible option for protecting the clinical data repository in CHIC, provided the repository satisfies the criteria for protection, i.e., shows significant investment in "*the obtaining, verification or presentation*" of its contents. The CHIC data repository hosts data categorized per data type: imaging data (DICOM etc), descriptive/structural data (age, sex etc), other files (histological reports), links (to other data repositories) etc.[108] The datasets for each type are accessible

[104] Ibid, WP6, T.6c), p.32.

[105] Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, OJEU, L 77/20 - L 77/28, 27.3.96.

[106] Recital 7 Database Directive.

[107] Article 1 Database Directive

[108] CHIC EC-GA, Annex I, DOW, WP6, Sub-Task 6.4c, p. 24.

individually so that the data corresponding to the model parameters may be chosen. The fact that the repository is built "*based on the experience already accumulated during the implementation of other data repositories*"[109] should be sufficient to prove the requisite investment in "*either the obtaining, verification or presentation"* of its contents. Against this background, the database right in the CHIC clinical data repository is likely to be granted.

As regards the scope of the database right, it would protect the collected data from being copied as a whole or in substantial part, evaluated "*qualitatively and/or quantitatively*"[110] and either copied in one action or step by step.[111] In this respect, the database right does not protect the data or isolated datasets per se; Article 7 (4) makes explicit that database right:

> "*shall apply irrespective of eligibility of the contents of that database for protection by copyright or by other rights. Protection of databases [….] shall be without prejudice to rights existing in respect of their contents*".

Thus, the database right may be an option to protect the data as a whole or in substantial part, but it will not extend to protect isolated data items. Therefore, if it is the datasets per se, or the investment deployed in processing the data, which are to be protected, the database right would not provide such protection.

6.2.2.2 Protection of data under copyright

As noted above, it may be the case that the database is protected by the database right, and the contents of that database – by other rights, such as copyrights. This, in particular, may be true for databases with some creative contents, such as music, videos, poems, etc. However, the law of copyright may hardly be considered as a plausible option for protecting clinical data, in particular when such data originate from the clinical trials or laboratory tests, where there is no creative input in it.[112]

The clinical data, collected in the CHIC repository, usually consists of some numeric parameters, figures, words, combinations of such items. However, the CJEU has clarified that isolated items, be they words, keywords, syntax, figures or mathematical concepts,will not attract copyright, such items, "*considered in isolation, are not as such an intellectual creation of the author who employs them. It is only through the choice, sequence and combination of those words that the author may express his creativity in an original manner and achieve a result which is an intellectual creation*."[113] Accordingly, the law of copyright would not present a suitable option for protecting the data, processed in CHIC.

6.2.2.3 Protection of data as know-how

A further means of affording protection to the CHIC data may be the rules dealing with confidential undisclosed information as 'know-how'. Such protection is provided by Section

[109] Ibid.

[110] Article 7 (1) Database Directive.

[111] Ibid, Article 7(5).

[112] The general requirements for copyright protection are discussed in deliverable D4.2.

[113] CJEU, Judgment of 16 July 2009, Case C 5/08, Infopaq International A/S v Danske Dagblades Forening, Ref. 45.

7, Article 39 et seq. TRIPS Agreement,[114] and enables natural and legal persons to prevent "*information lawfully within their control from being disclosed to, acquired by, or used by others without their consent in a manner contrary to honest commercial practices".*[115] Unfair practices for these purposes would include the acquisition of information via violation of contractual duties, breach of confidentiality obligations, inducement to breach, etc.[116]

In order to be protectable, the relevant information should have the quality of protectable information within the meaning of Article 39 TRIPS Agreement. Article 39 TRIPS Agreement protects information, which is:

> *"(a) is secret in the sense that it is not, as a body or in the precise configuration and assembly of its components, generally known among or readily accessible to persons within the circles that normally deal with the kind of information in question;*
>
> *(b) has commercial value because it is secret; and*
>
> *(c) has been subject to reasonable steps under the circumstances, by the person lawfully in control of the information, to keep it secret."*[117]

The first weak point of protecting the processed data as know-how is that as of now the legal framework on know-how protection in the EU is not harmonized. Although, there is a proposal for a draft directive on the protection of undisclosed information in the EU[118] (the Draft Directive), before it is adopted and implemented, protection of know how remains dispersed through the national states of the EU Member States, and subject to varying requirements for and scope of protection.[119]

The Draft Directive, which is intended to harmonize the national laws in relation to know-how protection, in many aspects repeats the provisions of the TRIPS Agreement (in particular, it relates to the protectable subject matter and requirements for protection (Article 2), acts of unfair acquisition of information (Article 3), rights and remedies conferred (Article 5 et seq), etc.). In this regard it may also be queried how far the Draft Directive, if adopted, would improve the protection for data, the preparation of which consumed much effort, but which for one or another reason may not reach the level of protectable know-how. Here the key obstacles in applying know-how protection to the processed data in research relate to the need (in order to be protected) for such data to be be secret, subject to the confidentiality measures and have economic value.

Thus in CHIC, as noted, there is a large mass of research data, comprising multiple data types, formats, words, figures, numerical parameters, abbreviations, etc. In order to satisfy the requirements for know-how protection, this data must first be secret and treated as

[114] Agreement on Trade-Related Aspects of Intellectual Property Rights, the TRIPS Agreement, Annex 1C of the Marrakesh Agreement Establishing the World Trade Organization, Marrakesh, Morocco, 15 April 1994.

[115] Article 39 (2) TRIPS Agreement.

[116] Article 39 TRIPS Agreement.

[117] Article 39 (2) TRIPS Agreement.

[118] Proposal for a Directive on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure, Brussels, 28.11.2013, COM(2013) 813 final, 2013/0402 (COD).

[119] Hogan Lovells International LLP, Report on Trade Secrets for the European Commission, Study on Trade Secrets and Parasitic Copying (Look-alikes), MARKT/2010/20/D.

such and, second, it must possess economic value. To satisfy the criterion of secrecy, the initial datasets must be stamped as “Confidential”. In relation to the volumes of research data in CHIC, this requirement, besides being at odds with the underlying culture of academic research, would create further workload. Moreover, this confidentiality mark would then need to be borne by the data throughout the whole multiscaling process so that data marked as “confidential” by the input into the model is marked “confidential” by the model output, and is stored as “confidential” in the in silico trial repository, etc. This preservation of the confidentiality marks would present another challenge.

Against this background, know-how protection might, in principle, be considered as a possibility for protecting some defined amount of data, but hardly makes for a feasible solution, when protection of all clinical data processed in CHIC is sought. It also may operate against the principle of openness desirable if optimal use is to be made of the data by the health research community, exploiting the full potential of skilfuly curated data sets.

6.2.2.4 Protection of investment in producing the data by related rights

Apart from the rights considered so far, there are in the field of intellectual property law a number of other emerging rights granted as a response to relevant investment. These rights are normally provided to the person, who invests in producing the protectable information. Such rights are referred to as related rights. Protection by related rights does not necessarily link to the intellectual creation (as the case is with traditional intellectual property rights, such as copyrights or patents), but rather to the economic investment.

The major rationale for protection by related rights tends to shift between intellectual creation and investment.[120] A mixture of artistic creation and investment attracts exclusive rights to performers in fixations of their performances. The economic investment constitutes a major factor, which renders exclusive rights to phonogram producers in their phonograms, to the film producers in respect of first fixations of their films, to broadcasting organizations in fixations of their broadcasts[121]. However, the number of related rights as of now is rather limited (mostly to those, indicated above), and does not apply to the digitalized data.

6.2.3 Ways forward for protecting investment in data curation

As has become apparent, data collection and curation efforts are a critical part of producing robust digitalized data for use in E-health research, and later informing clinical practice. However, quite aside from the privacy risks to patients and research subjects analysed in Chapter 4, a further significant impediment to faster progress in E-health research is the failure of existing schemes of IP law adequately to recognize and reward the investment of time and skill it takes to acquire reliable, good quality data. As will be discussed later, this issue is now also attracting academic attention and initiatives at EU level. [122]

[120] H. Zech, Information als Schutzgegenstand, at p. 142.

[121] DIRECTIVE 2001/29/EC of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society, OJEU L 167/10 - L 167/19, 22.6.2001.

[122] The EC is also planning to engage in consultation on copyrightability/ownership of data: see https://cdt.org/files/2015/12/Communication-towards-modern-more-European-copyright-framework.pdf] and [https://cdt.org/blog/european-commission-copyright-action-plan-a-busy-2016/].

At present, the tendency is for researchers and institutions that engage in such efforts to protect their investment by keeping the data confidential, and releasing it to other researchers – if at all – on license terms that seek to give them a future return on profitable uses made with the data. The disadvantage of this 'data-siloing' approach[123] is that it leads to a fragmented and secretive research environment, inimical to the adoption of transparent and consistent curation standards, required for greater data interoperability to be achieved. In this part, we briefly consider possible approaches from the academic legal literature that might point to a way forward.

First, an important aspect is to define the terms "information" and "data" more closely, as part of determining what the putative object of legal protection should be. This would serve as an important clarificatory step, as currently different areas of science operate with diverse terms. Also, the various fields of law operate different terms. Thus, when it comes to personal data, the law on data protection operates with the term "data".[124] The criminal law, dealing with information in the context of computer technologies, speaks about "computer data", meaning "*any representation of facts, information or concepts in a form suitable for processing in a computer system, including a program suitable to cause a computer system to perform a function*"[125]. By contrast, intellectual property law and the competition law, when protecting know how, use the term information[126]. Dreier suggests understanding the "*information in the sense closer to "data" than to knowledge*"[127] and highlights the role of IP rights as "*legal rules which [...] regulate the generation, flow, storage and use of information*".

Here one option might be to create property rights in information per se. This has been extensively discussed in the legal literature without a clear consensus being reached. For example, among German commentators, Wagner suggests a right on binary information, generated in the result of recording a person, such as by computer tomography, taking photo of a person by digital camera, or recording a speech, etc.[128] This approach may be arguable, either from perspective of the law on data protection (sometimes lacking relation to an identifiable data subject), or from perspective of the intellectual property law, which grants protection in reward for intellectual creations or investments.[129] Similarly, Peukert discusses "informational goods", as protectable subject matter, and seeks to find points for relating such goods, which are not covered either by the ownership in property or by the intellectual property rights, to the right holders.[130]

This focusing on informational goods as object of protection would allow developing a system of attributable legal rights and factors, which could be taken in consideration when

[123] European Commission 'Redesigning health in Europe for 2020', p. 10.

[124] Article 2 Data Protection Directive.

[125] Council of Europe, CONVENTION ON CYBERCRIME, Budapest, 23.XI.2001.

[126] Article 39 TRIPS Agreement; Commission Regulation (EU) No 316/2014 of 21 March 2014 on the application of Article 101(3) of the Treaty on the Functioning of the European Union to categories of technology transfer agreements Text with EEA relevance OJ L 93, 28.3.2014, p. 17–23.

[127] T. Dreier, 'Regulating information: Some thoughts on a perhaps not quite so new way of looking at intellectual property'.

[128] A. Wagner, Binäre Information als Gegenstand des Rechtsverkehrs.

[129] H. Zech, supra, p.297.

[130] A. Peukert, Güterzuordnung und Freiheitsschutz.

attributing such rights. At the same time, as noted by other writers, there are also some problematic issues with respect to attributing exclusive rights in information in the context of intellectual property law.[131] Granting IP rights in something as basic as information might be analogized to creating property rights in the air we breathe and is particularly problematic in light of the grounds for justifying such protection. Thus it may operate against the free flow and exchange of ideas essential in an open society. In the context of personal data or information, there is also the point that the data subject might reasonably be thought to have the greatest prima facie property claim; however, this would clearly not be a basis for rewarding the investment of third party curators; indeed, in the health research context, with the important good of medical progress as its aim, to assign property rights to the subject rather than the researchers who seek to find valuable uses for the data could be seen as a retrograde step.[132]

Here, a more differentiated approach has been suggested by Zech, who bases his analysis on the division of information into three layers: semantic, syntactic and structural. This division resides on differentiation of content layer, code layer and physical layer, represented by Benkler[133] and Lessig[134]. The semantic layer of Zech corresponds to the contents, which the information bears, syntactic layer is the level of symbols, in which such information is recorded, the structural layer would be the tangible medium, where such information is embodied. Zech considers that exclusive attribution of information is possible not only in the context of intellectual property rights, but also in the other areas of law, in particular, by personal rights and the property law. The primary goal here is to categorize the existing legal rights in relation to the various informational goods[135].

Building on this, a starting point in the context of curation might be to see raw data in terms of the 'given', which itself lacks semantic meaning, with the latter only emerging through the addition of an interpretive context (which also marks the change in state from data into information). It is arguably the technological development and transformation of raw or incompletely processed data into information (or the uncovering of additional semantic meaning), brought about by the curative process, which presents the suitable object of IP protection.

As noted, the present approach that seeks to maintain (commercial) data confidentiality by silo-ing data leads to a fragmented research environment, and reduces the chances for greater data interoperability to be achieved. Here the law - aided by technology should aim to encourage greater openness, while assuring appropriate curation rewards. This could, e.g. take the form of an officially endorsed mechanism or system for measuring and tagging changes produced in a given data set (or the merging of several data sets) resulting from curation efforts, as the reward-trigger. At the same time, as another crucial policy element, the law needs – especially in the case of the curation of sensitive health data – to ensure

[131] R. Weber, Ali Baba oder das Risiko exklusiver Informationsinhaltsrechte.

[132] See by analogy the US decision of Moore v Regents of University of California 73 Pd 479 (1990), denying a patient property in his extracted tissue later used by researchers to generate a lucrative cell-line.

[133] Y. Benkler, From Consumers to Users: Shifting the Deeper Structures of Regulation.

[134] L. Lessig, The Future of Ideas, The Fate of the Commons in a Connected World, New York 2002.

[135] H. Zech, supra, p. 2.

that, as discussed in Chapter 4, privacy and other interests of patients and research subjects are and remain adequately protected.

In particular, it will here be necessary to take account of (and compensate for) the knock-on effects of IP changes, where data-holders are no longer (also) motivated by commercial considerations to keep their data secure and confidential. This concern is all the greater here since the activities of data sharing and curation being encouraged, also by their nature present enhanced risks to personal privacy. The point of curation is precisely to uncover new connections and patterns in data that help generate robust inferences (usable – for good or ill) about the relevant data subjects. Accordingly, it is submitted that any system for rewarding investment in data curation should also require (as a condition for such rewards) that the data curator takes every appropriate measure to counterbalance the associated enhanced risks to privacy. Here, a model under which curated data is sent to secure publicly stewarded repositories, as under development in the EUDAT project,[136] may suggest itself. In any case, the Commission's plan to consult with and solicit feedback from stakeholders on these issues is to be welcomed.[137]

6.3 Rewarding research activity in cancer modeling

6.3.1 Modelling and almagamating models in CHIC

As described under 6.1, the process of multiscale cancer modelling such as occurs in CHIC is divisible into several stages. First, scientific modeling work is performed, in which the modelers study the tumor types and biological processes, selected for simulation, investigate the types of cells and interactions among them, break down these processes into the elementary biological processes (biomechanics), such as cell cycling, the angiogenesis process, declination of a cell to apoptosis after a particular treatment, etc. Secondly, work is done to transform those scientific models into computer models, which are encoded into computer programs. In this process, either the already developed codes of tumor models are broken down into simpler models or computer codes of elementary biological processes (biomechanics) are developed anew. Coding finalizes the development of cancer models.[138]

Thirdly, the simulation of cancer progression in space and time requires the use of multiscale cancer modelling.Multiscaling is realized in silico by constructing elementary models (the ones which correspond to elementary biological processes) and relation models (the ones which reflect relations across them) into multiscale hyper-models. "*A model is considered to be "multiscale" if it spans two or more different spatial scales and/or includes processes that occur at two or more temporal scales*."[139] The four main biological scales which are being modelled are the atomic, molecular, microscopic and macroscopic scales.[140]

In CHIC, research groups from the different project partner institutions contribute single-scale models (from molecular to compartment models), which are then combined into

[136] [http://www.eudat.eu/].

[137] See: [https://cdt.org/blog/european-commission-copyright-action-plan-a-busy-2016/].

[138] CHIC Grant Agreement no: 600841, Annex I - „Description of Work", WP2. WP6.

[139] T. Deisboeck, et al, 'Multiscale Cancer Modelling'.

[140] Ibid.

integrated multiscale hyper-models. Processes which occur at the atomic level are linked to the processes at a higher level. The composite multiscale constructs of models (hyper-models or integrative models) are then able to synthesize and simulate the biological processes at several different levels (molecular, cellular, etc.) at once so that spatiotemporal simulation of the clinical trials and studies, addressed by the CHIC project, is achieved.[141]

The hyper-modeling execution process itself is to a large extent facilitated and semi-automated by the underlying technical infrastructure (e.g. in CHIC, hyper-modeling is managed by VPH-Hypermodelling Framework[142]). However, at the earlier stage of hyper-model construction, human input into the hyper-model design is indispensable. Particularly where the intention is to use hyper-models for decision support in the clinical setting, the automatic linking of hypo-models is unjustifiably risky and thus not acceptable[143]. The hypermodeling strategy, the hypomodel linking, the hypomodel integration are instead normally designed by the modeling party, who has substantial expertise in the field of bioinformatics. As discussed below, it is this intellectual input, deployed in designing a hyper-model, which may deserve protection and make a hyper-model into a work, protectable under the IP regimes of patent or copyright law. We consider these regimes in turn.

6.3.2 Potential IP protection for models under patent law

6.3.2.1 Basic requirements and exclusions

The legal framework of the patent law, including the nature of patentable subject matter and the patenting requirements, have been described in the CHIC Deliverable D.4.2, Section 5.4,[144] without the need to repeat these general considerations in detail here. As noted there, the basic principle of the patent law, namely that *"patents shall be available for any inventions, whether products or processes, in all fields of technology, provided that they are new, involve an inventive step and are capable of industrial application".*[145] These requirements are followed in a precise and strict manner by patent tribunals and courts when determining applications for protection.

Thus, according to the main principles of patent law[146], to be patentable an invention must be novel, expose an inventive step and be susceptible of industrial application. The requirement of novelty would be fulfilled if substance of the invention has not be disclosed to the public before the application is filed[147]. Disclosure to the public (thereby depriving the invention of protection) may occur by any means, such as: publications, performance to the audience at exhibitions, workshops, etc., patent applications, etc. Secondly, an invention

[141] CHIC Grant Agreement no: 600841, Annex I - „Description of Work", WP6.

[142] D. Tartarini, et al, 'The VPH Hypermodelling Framework for Cancer Multiscale Models in the Clinical Practice'.

[143] G. Stamatakos et al, 'The Technologically Integrated Oncosimulator: Combining Multiscale Cancer Modeling with Information Technology in the In Silico Oncology Context'.

[144] CHIC, Deliverable No. 4.2, Initial analysis of the copyright-related legal requirements for the sharing of data.

[145] Article 27 (1) TRIPS Agreement.

[146] Article 52 (1) European Patent Protection (EPC), Article 27 (1) TRIPS Agreement.

[147] Article 54 EPC.

exposes an inventive step, if the technical progress which it achieves is not obvious to the person skill in the art.[148]

The third requirement is that an invention is susceptible to industrial application, and means it is capable of being used in any industry.[149] However, it also links to a fourth, implicit patenting requirement, which is that of technicity. in this regard, the patentable invention must have a "technical character", i.e. involve a "technical teaching", meaning instruction to a "*technically skilled person as to how to solve a particular technical problem using particular technical means*."[150] It follows that a problem solved by the invention, must be technical, in contrast to purely commercial or mathematical one.

However, a complicating factor is that, by way of derogation from the above framework, certain areas of subject matter have expressly been excluded from patentability by the European Patent Convention (EPC). In particular, under Article 52(2) EPC, such excluded matter comprises (a) discoveries, scientific theories and mathematical methods; (b) aesthetic creations; (c) schemes, rules and methods for performing mental acts, playing games or doing business, and programs for computers; and (d) presentations of information.

Here the exclusion related to computer programs has led to much not always consistent case law, with the dividing line between what may still be patentable (a program that performs an external technical operation, as opposed to 'software as such') still obscure. As one leading textbook in the area notes, "*there is no bright line between what is, or what is not, within the exclusions [of software patentability] and the border has proved difficult to delineate*".[151]

Similar uncertainty arises from the exclusion under Article 53(c) EPC of "*methods for treatment of the human or animal body by surgery or therapy and diagnostic methods practised on the human or animal body*", followed by the proviso that "this provision shall not apply to products, in particular substances or compositions, for use in any of these methods." In one decision, it was suggested the application of the exclusion would depend on whether a given product's design presupposed a prior therapeutic or diagnostic intervention. Thus, whereas "*a method of manufacturing an artificial limb should be patentable [as] a moulding of the stump on which an artificial limb is fitted is clearly not of a surgical nature…or require the presence of a medically qualified person…, a method of manufacturing an endoprosthesis outside the body, but requiring a surgical step to be carried out for taking measurements, would be excluded from patentability*".[152]

6.3.2.2 Application of patenting requirements to cancer models

The main points, which may be made when considering patenting of the hyper-models and assessing the probabilities to obtain the patent rights in models, are the following. First, in principle, given the innovative substance of the cancer models and the potential of their

[148] Article 56 EPC.

[149] Article 57 EPC.

[150] EPO, Patents for Software, http://www.epo.org/news-issues/issues/software.html, retrieved 04.03.2016.

[151] D. Rowland et al, *Information Technology Law*, Routledge, 4th Ed., 2012, p. 377.

[152] T 1005/98.

application as a decision support tool in clinical practice,[153] obtaining rights in some hyper-models under patent law may appear prima facie possible.

Thus, as established above, a cancer model in CHIC is defined as: "*a computer program that implements a scientific model, so that when executed according to a given set of control instructions (control inputs) computes certain quantities (data outputs) on the basis of a set of initial quantities (data inputs), and a set of execution logs (control outputs).*"[154] By way of providing a concrete illustration, let us take a hyper-model consisting of two component models: a microscopic cellular bio-model, which shows the concentration of cells in the invaded tissue, and a macroscopic biomechanical model, which provides directions of least pressure in the tissue.

Here the model of tumor growth is composed from a macroscopic mechanical model, which provides directions of least pressure in the tissue, and the cellular level model, which simulates concentration of cells inside the element. The microscopic cell-level model requires the direction of cells proliferation by tumor growth as an input. This information is provided by the mechanical model. In its turn, the biomechanical model requires the number of cells inside the element. This information is calculated by the biological cell-level model and fed to the biomechanical model. This model of tumor growth is structured via interplay and exchange of information between the models. The overall function of the hyper-model is to simulate the "*geometrical evolution of the tumor predicted at the cellular level*."[155]

This method of producing a cancer model simulating a tumor growth has a potential to be applied in clinical practice to predict evolution of the tumor. By this token, the patenting requirement of industrial application may be estimated as achieved. However, the difficulties in satisfying the other requirements remain significant. Thus, as regards the need for novelty, this may not be fulfilled, because the process of generating this model has been previously disclosed to the public, both at the cancer modeling workshop and in the scientific publication.[156] Similarly, state of the art knowledge for building hypermodels, which is also used in hyper-modelling in CHIC, has also been disclosed in the CHIC public deliverable D.2.1[157] The technical character of a problem, solved by the model, may also be questioned, since the method of combining this model is rather to be used to assist the clinical in making a decision what method of therapy to apply, than solves some "*particular technical problem using particular technical means*".[158]

As already flagged, a further key difficulty in seeking patent protection for in silico models, at least in Europe, concerns the various patent exclusions under the EPC, such as the fact that computer software is not patentable "*as such*".[159] This would mean that a cancer model "as such", which constitutes computer software, is excluded from patenting under Article 52

[153] F. Rikhtegar, et al, 'A Model of Tumor Growth Coupling a Cellular Biomodel with Biomechanical Simulations' (2014).

[154] CHIC, Deliverable D7.1, 'Hypermodelling specifications'.

[155] F. Rikhtegar, at al, supra.

[156] Ibid.

[157] CHIC, deliverable D2-1, State of the art knowledge for building hypermodels.

[158] EPO, Patents for Software, http://www.epo.org/news-issues/issues/software.html, retrieved 04.03.2016.

[159] Article 52 (2) c EPC.

(2)(c) EPC. Admittedly, it may be possible to argue instead that the models/hyper-models in CHIC in fact qualify models as computer implemented invention (CII), making them patentable after all. As noted, though, this line may not be straightforward to draw.

In this context, patent protection for a method for processing medical data for patient examinations with the use of artificial intelligence was denied in Germany on the ground of non-patentability of rules and methods for performing mental acts and presentations of information[160]. The claimed subject matter contained instructions related to the choice of examination modalities (e.g. X-ray, computer tomography, magnet resonance) and purpose-related application on the patient by means of a program using a symptom-specific and/or diagnosis specific database.

The patent tribunal found that instructions on selecting one or more examination and measurement protocols as well as selection of examination modalities by a physician constitute non-technical elements which only aim to automate decisions already contemplated by the physician and solve no technical problem at hand. On that basis the claimed method was declared non-patentable as such[161]. For its part, the appeal court, while suggesting that in principle a method that is using a computer program to solve a technical problem could be patentable, agreed that here a technical problem which the claimed method solved could not be identified. Accordingly, patent protection was denied.

At the same time, the case law on patenting CII is constantly evolving and possibly a more relaxed approach to patenting software will be adopted in the future. This could also bring the European approach into line with the more relaxed requirements for patenting CII under US law, "*where patent protection for software is granted, even if it does not solve a technical problem*".[162] What may also be considered potentially patentable in CHIC is the process of combining models into a hyper-model. In the case of a model of tumor growth described above, where a cellular bio-model is coupled to biomechanical simulations,[163] the process of producing such model and steps describing the sequence and manner of combining such component models, might define the substance and be brought forward as a claim of a patentable invention.

At the same time, in the light of the current exclusions and uncertainties in this area of law, we would not recommend this approach to partners in the lifetime of the project. Indeed the disincentives for in silico modelers to consider patenting their inventions in Europe are accentuated by the high potential financial costs of applying for such protection – requiring both the services of a specialized patent attorney – and the registration fee payable to the patent office. In this regard, the alternative IP regime of copyright, which arises automatically without any need for registration, may appear more flexible and attractive as a means to protect modeler investment. We consider the application of copyright in the section below.

[160] Case Law from the Contracting States to the EPC 2004 – 2011, Special edition 3, Official Journal EPO 2011, Federal Court of Justice of Germany, Judgment of 20 January 2009 - X ZB 22/07 - Equipment for selecting medical examination methods.

[161] Federal Court of Justice of Germany, Judgment of 20 January 2009 - X ZB 22/0, Rn.5.

[162] EPO, Patents for Software, http://www.epo.org/news-issues/issues/software.html, retrieved 04.03.2016. The patentability of cancer models in CHIC under the US law is not analyzed in this report.

[163] F. Rikhtegar, at al, supra.

6.3.3 Potential IP protection for modellers under copyright law

6.3.3.1 Copyright in computer software

Copyright is a traditional form of IP protection, which in more recent times both European and International law have extended also to computer programs. Article 4 WIPO Copyright Treaty[164] and Article 10 TRIPS Agreement[165] protect computer programs as literary works within the meaning of the Berne Convention (1886)[166]. The same principle is followed by European copyright law. Article 1 of Directive 2009/24/EC on the legal protection of computer programs (Software Directive 'SWD'))[167] recognizes computer programs as an object of copyright protection in the EU. Here Article 1 (2) grants copyright protection to "*the expression in any form of a computer program*".

What counts as a program expression for the purposes of the SWD has been established by the CJEU in its case law. Thus in the case Bezpečnostní softwarová asociace – Svaz softwarové ochrany (BSA)[168] the court, in consideration of the international copyright law, and Article 10 TRIPS Agreement, held that the source code and the object code of a computer program are forms of expression entitled to be protected by copyright. As it stated, "*the object of the protection conferred by that directive is the expression in any form of a computer program which permits reproduction in different computer languages, such as the source code and the object code*."[169] The source code of a program usually constitutes a script, written in a human readable form. The source code, compiled into a binary executable, constitutes the object code, which gives the final instructions to the computer[170].

The SWD protects programs, which are "*original in the sense that it is the author's own intellectual creation*."[171] Original intellectual creation is a basic requirement for copyright protection. It is equally applicable to the other copyright works, be it software, writings, photographic works, or other works protectable by copyright; no other criteria, such as whether a program is functional, or how many lines of code it has, etc. are relevant. At the same time, the requirements of originality and expression have a specific interpretation in terms of copyright in software. This reflects the fact that, compared to the use of ordinary language (as where a literary author arranges words in a sentence), computer code has special features that argue in some instances for a narrowing of the copyright in a particular sequence of code, and in other case for a widening of copyright, so that it subsists even where precise sequences of code are not copied.

[164] WIPO Copyright Treaty, Geneva, 20 December 1996.

[165] Agreement on Trade-Related Aspects of Intellectual Property Rights, the TRIPS Agreement, Annex 1C of the Marrakesh Agreement Establishing the World Trade Organization, Marrakesh, Morocco, 15 April 1994.

[166] Berne Convention for the Protection of Literary and Artistic Works of September 9, 1886.

[167] Directive 2009/24/EC of the European Parliament and of the Council of 23 April 2009 on the legal protection of computer programs, Official Journal of the European Union, L 111/16 – 111/22, 5 May 2009.

[168] CJEU, Judgment of 22 December 2010, Case C 393/09, Bezpečnostní softwarová asociace – Svaz softwarové ochrany v Ministerstvo kultury.

[169] CJEU, BSA, supra.

[170] U.S. Court of Appeals, Third Circuit, Ruling of August 4, 1986, Whelan Associates Inc. v. Jaslow Dental Laboratory, Inc., et al, 797 F.2d 1222, 230 USPQ 481.

[171] Article 1 (3) SWD.

As regards the first 'narrowing' aspect, the courts have been required to take account of the fact that certain sequences of syntactical code may be indispensable for the overall code to function. Here it would be undersirable to recognize copyright, as this would confer a monopoly on the first programmer who came up with the given sequence. Accordingly, in the case SAS Institute Inc.,[172] the court concluded "*neither the functionality of a computer program nor the programming language and the format of data files used in a computer program in order to exploit certain of its functions constitute a form of expression of that program for the purposes of Article 1(2) of Directive 91/250*." The CJEU supported its decision by the argument that "*… to accept that the functionality of a computer program can be protected by copyright would amount to making it possible to monopolise ideas, to the detriment of technological progress and industrial development*." [173] Instead it is only where the programmer "*by the choice, sequence and combination of those words, figures or mathematical concepts*" in his program succeeds in expressing his creativity in an original manner, that this justifies the protection of a program by copyright[174].

Conversely, and relevant to the 'widening' aspect, there are other non-literal elements, used in computer programming, which define perception of a program by a user and constitute its non-literal expression. In this respect it may be possible for new software to mimic very closely the 'look and feel' of another given software program without using the same coding sequences at all: here, the question on protection of non-literal expression of a program by software copyright has been raised for consideration. In response, courts, especially in common law countries, such as the UK and US, have tended to extend the scope of software copyright beyond the literal code to a program non-literal expression.[175] The UK courts approach copyrightability of non-literal expression as follows: "*Consideration is not restricted to the text of the code… That must be right: most literal copyright works involve both literal matter (the exact words of a novel or computer program) and varying levels of abstraction (plot, more or less detailed of a novel, general structure of a computer program)*."[176]

In this regard, the issue of how far a programmer's structuring efforts in planning and executing a software program should similarly be protected against 'non-literal copying' (in the face of the traditional 'ideas-expression dichotomy', where 'structuring' might be seen as part of the non-protectable realm of ideas) is of special salience. Thus, in practice It is often the case that structure of one program is imitated or reproduced by another program and it when copyright protection for a program structure is sought. The legal issue raised before the courts here is: "*whether the structure (or sequence and organization) of a computer program is protectible by copyright, or whether the protection of the copyright law extends only as far as the literal computer code*".[177]

[172] CJEU, Judgment of 2 May 2012, Case C 406/10, SAS Institute Inc v World Programming Ltd; see also CJEU, Judgment of 16 July 2009, Case C 5/08, Infopaq International A/S v Danske Dagblades Forening.

[173] CJEU, SAS Institute, supra, Ref.40.

[174] CJEU, SAS Institute, supra.

[175] S. Stokes, "Digital Copyright, Law and Practice", 4. Edition, Hart Publishing, Oxford and Portland, Oregon, 2014, p. 121.

[176] Ibcos Computers Ltd v Barclays Mercantile Highland Finance Ltd [1994] FSR 275.

[177] Oracle America, Inc., v. Google Inc., Case C 10-03561 WHA, supra.

At present, in Europe neither the SWD affords such protection, nor has the CJEU addressed copyright protection of non-literal program expression in its case law. However, the reason why the issue is important is that structuring a program often takes more time and intellectual effort then writing the code. The code of a program itself consists of a set of instructions to the computer and is an end product of a complex software development process[178]. The latter process often occurs in several steps. The first step in software development is identifying the problem that the computer programmer is trying to solve. In the next step, the outline for the solution follows. In the outline, the programmer breaks down the solution into smaller units called 'subroutines' or 'modules', each of which handles elements of a larger problem.

The outline can be laid down in the form of a flowchart. The next step is organizing the modules and subroutines into a program structure. A program structure is dictated by "*the functions of the modules in a program together with each module's relationships to other modules*"[179]. Usually, the modules are then arranged so that the problem is solved in a more efficient way. As observed by the court in the US case of Whelan Associates Inc. v. Jaslow Dental Laboratory:[180] *"Although two programs could produce the same result, one might be more efficient because of different internal arrangements of modules and subroutines. Because efficiency is a prime concern in computer programs (an efficient program being obviously more valuable than a comparatively inefficient one), the arrangement of modules and subroutines is a critical factor for any programmer."*

After defining the structure, a programmer decides about what data is needed, where and how the data should be introduced and how it should be combined with other data. The data arrangement is achieved via data files. Once the program design is ready, the coding begins. The coding is a comparatively small part of programming. "*By far the larger portion of the expense and difficulty in creating computer programs is attributable to the development of the structure and logic of the program, and to debugging, documentation and maintenance, rather than to the coding*."[181]

Here, an underlying challenge is to separate those aspects of the programmer's overall efforts, for which his individual creative contribution merits protection by copyright, from other aspects that should remain outside such protection, either for reasons of policy (as general ideas that should be left in the public domain for others to use freely), or simply because they fail to evince sufficient originality. An influential judicial response in the US to this is the so-called "abstraction-filtration-comparison" test, established in the case of Computer Associates International, Inc. v. Altai, Inc.[182]

[178] Report by the International Bureau, "Measures to enhance International Cooperation in the field of Legal Protection of Computer Software", WIPO expert group on the legal protection of computer software, First Session, Geneva, 27-30 November 1979, LPCS/I/2, 30 September 1979.

[179] Oracle America, Inc., v. Google Inc., Case C 10-03561 WHA, supra.

[180] U.S. Court of Appeals, Third Circuit, Ruling of August 4, 1986, Whelan Associates Inc. v. Jaslow Dental Laboratory, Inc., et al, 797 F.2d 1222, 230 USPQ 481.

[181] Whelan Associates, ibid.

[182] U.S. Court of Appeals for the Second Circuit, Ruling of 17 December 1992, Computer Associates International, Inc. v. Altai, Inc., 982 F.2d 693, 2d Cir.1992.

The test comprises three steps. First, a copyright program is broken down into its structural components according to the levels of abstraction. The second step extracts certain non-copyrightable structures (discussed below) until the copyrightable substance remains. In the third step, what has been copied is compared with the copyrightable expression in a structure of the original program. Finally, it is estimated in how far such copying is substantial relative to the overall program[183].

The three types of structures, identified as precluded from copyright, comprise: (i) elements dictated by efficiency, (ii) elements required by external factors, (iii) elements taken from the public domain. As regards the first, copyright does not extend to structures dictated by efficiency (the doctrine of merger). Accordingly, copyright will not subsist in the expression, which is "*necessarily incidental to the idea being expressed*." In this step, it is determined "*whether the use of this particular set of modules is necessary efficiently to implement that part of the program's process" being implemented.*" If efficiency dictates that the choice of modules is limited to just a few workable solutions (such as one or two options), such selection of modules may not be protectable as such.[184] .

Secondly, copyright does not extend to structures dictated by external factors. External factors in software programming may include: compatibility requirements, mechanical specifications, computer manufacturer design standards, industry demands, and common programming practices. In US copyright law, this is known as the 'scenes a faire' doctrine. In consequence, a particular set of modules, which need to be present as an integral part in all programs of the same category are non-copyrightable. Thirdly, copyright does not protect structures that are found in the public domain. The rationale here is that such material, which is included as an element in a copyrightable work, may itself not be appropriated as it should remain free for use by the community[185].

As discussed under 6.3.3.4 below, the issue of copyright in structure also resonates in the field of cancer modeling, particularly at the stage when individual models are amalgamated into hypermodels.

6.3.3.2 Potential application to the models in CHIC

As we have seen above, the cancer models, encoded into computer programs, would constitute subject matter protectable by copyright, provided they satisfy the requirements for such protection. Here the model must amount to an original expression in a form, which counts as a program expression for the purposes of software copyright.[186] According to the case law of the CJEU, such expression is present, when the modeller "through the choice, sequence and combination" of commands in the model code succeeds to "*express his creativity in an original manner and achieve a result which is an intellectual creation*".[187]

Here it is the code, in which a model is embodied, be it the source code written in Python or the code compiled in C++, which is protected by copyright. The programming language itself,

[183] Altai, Ibid; see also the discussion in Stokes, note 153 above.

[184] Altai, Ibid.

[185] Ibid.; a more recent case, illustrating the application of the test, is Oracle America, Inc. v. Google Inc., C 10-03561 WHA.

[186] Article 10 TRIPS Agreement

[187] CJEU, Infopaq, supra.

the biological process implemented by the model, the general process of its implementation, both as formats of data files, used by exchange of data between the models remain outside the scope of copyright protection. Similarly, if the code is generated automatically, for example by automatic translation into another programming language, or compiled from bits of code copied from the public domain, copyright protection in such code is most likely to be denied.

However, these are not only the codes of computer models themselves, which may enjoy such protection, also, the scientific modelling work as well as designing models into hyper-models may be copyrightable. In the next section, we explore how far copyright, as a type of protection applicable to the cancer modelling, may extend backwards – so to speak – so as to afford protection to earlier aspects of the modeller's intellectual investment in developing the scientific cancer models as the necessary precursor of the coded computer models. Following that, in section 6.3.3.4, we then turn to address the subsequent protection by copyright of the hypermodels that are created as amalgams of the constituent models. Here we are concerned with the conditions for a hypermodel itself as a composite entity, to attract copyright (over and above the separate copyrights that inhere in the discrete constituent models).

6.3.3.3 Copyright in preparatory design material

As noted earlier, a computer program stands at the end of a long development process, in which the model code, embodied in the computer program results from the foregoing modelling work. A code of a program may be compared with the tip of the iceberg, which reaches out above the water, where 90 per cent is hidden under the surface. And the same may be said in relation to the earlier scientific modeling (which precedes the computer modeling). In view of the skill, time and labour, which a programmer invests in studying a problem, elaborating a solution for it and making the solution executable by a computer, a justifiable legal question arises whether this pre-programming work is protectable by software copyright.

In fact, in contrast to some of the other issues considered in this Chapter, here the SWD gives a reasonably clear affirmative answer. It does so by extending the protection given by software copyright to the preparatory design work that preceded the creation of a computer program. In particular, Article 1 (1) of the SWD, when defining the object of protection, provides: "*the term 'computer programs' shall include their preparatory design material*." What is meant by the "preparatory design material" is explained in Recital 7 in terms of preparatory design work of a kind "*such that a computer program can result from it at a later stage*." As further interpreted by the CJEU, the preparatory design work counts as protectable by the directive along with a program, if it is "*capable of leading, respectively, to the reproduction or the subsequent creation of such a program*"[188].

By contrast to the form of program expression, protectable by copyright, there are no specific requirements to the form or mode in which the preparatory design work must be expressed. According to one decision of the Federal Court of Justice of Germany, which dealt with protection of preparatory design work by copyright, it will be sufficient that the development documentation was recorded in writing, such as: data flow plans, designs of

[188] CJEU, BSA, supra.

commands and information cycles, exhibits of scientific or technical art, expressed in any form, including mathematical, or technical or graphic symbols.[189] It follows, in relation to copyright in the kind of work occurring in CHIC, that modeling work (a) documented in writing, such as laid down in flow charts, exhibits, etc, and (b) attributable to a specific model, which if necessary may be rewritten from these materials, has a good chance of being protected by software copyright along with the model code.

6.3.3.4 Copyright in hyper-models as amalgamations of constituent models

As we saw earlier, multiscale cancer modeling is ultimately achieved by combining simple models into hyper-models. The aim thereby is that in a given instance (e.g. when populated with the data for a given patient) the hyper-model is able to reproduce complex biological processes, involving multiple phenomena which are respectively captured in single models; it "*emerges from the composition and orchestration of multiple hypomodels, each one of which is capable of simulating a specific entity or phenomenon… and can simulate an entity or phenomenon that may be more complex than the ones simulated by each separate simpler model*."[190]

In this regard, the appropriate concept for qualifying hyper-models in terms of copyright is arguably that of a "compilation". Compilations, as works protectable by copyright, are introduced by Article 5 WIPO Copyright Treaty, Article 10 (2) TRIPS Agreement. The TRIPS Agreement provides: "*Compilations, whether in machine readable or other form, which by reason of the selection or arrangement of their contents constitute intellectual creations shall be protected as such*."[191] In the case of CHIC hyper-models, such intellectual creation may reside in the innovative "*composition and orchestration of multiple hypomodels*", where (as per the Infopaq decision) the modeller "*through the choice, sequence and combination*" of models in a hyper-model "*may express his creativity in an original manner and achieve a result which is an intellectual creation*."[192]

In the CJEU case law, the cases on infringement of copyright in compilations are mostly tried in relation to database rights. The reason is that the Directive 96/9/EC on the legal protection of database (Database Directive) also uses the term "compilation", referring to collections of works, data or other materials[193]. In this case, copyright is a reward for the intellectual effort, which the author deployed in his selection or arrangement of the contents in a database. In this regard, the CJEU in the case Football Dataco Ltd. and Others v Yahoo! UK Ltd. and Others[194] measured originality by reference to the author's having made free and creative choices, thereby stamping his work with a personal touch. On the other hand, if external technical constraints leave no space for creative choice, the element of originality in the structure will be lacking.

[189] Federal Court of Justice of Germany, Judgment of 09 May 1985, Case I ZR 52/83, BGHZ 94, pp. 276 – 292.

[190] CHIC, Deliverable No. 7.1, "Hypermodelling specifications", CHIC, Downloads, Deliverables, Public Deliverables, <http://chic-vph.eu/uploads/media/D7-1_Hypermodelling_specifications.pdf> 2016-02-07.

[191] Article 10 (2) TRIPS Agreement.

[192] CJEU, Infopaq, supra.

[193] Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, OJEU, Nr. L 77/20-77/28, 27.3.96.

[194] CJEU, Judgment of 1March 2012, Case C 604/10 Football Dataco Ltd. and Others v Yahoo! UK Ltd. and Others.

Hyper-models, which are composed of single models, arranged in specific relations to each other, may also be protectable by copyright as compilations. For this, a hyper-model must show an original and creative structure. Copyrightability of a hyper-model as a compilation is to a large extent dictated and can be measured by the degree of originality and creativity, which a modeler deployed (and was able to deploy) in structuring his hyper-model. Thus, in cases where a hyper-model is structured in a particular way because there are no other ways of arranging models into it (such as when the model arrangement is pre-determined by technical requirements), it will not attract copyright. The same goes for cases where the structure relies on elements taken from the public domain. In contrast, if the modeler was not so restricted and expressed creativity stamping a hyper-model structure "with his personal touch", such investment shall not remain unrewarded and copyright in his work will likely be recognized.

If we compare the criteria for measuring creativity in a program structure, considered in 6.3.3.1, with the criteria of originality in the compilation, we can see that these criteria inter-relate with each other. Thus, if the author (either a programmer or a compiler of a database) had some scope to express his creativity in arranging the modules in a program (or contents into a database) by making free and creative choices, such a structure may show originality and qualify as an intellectual expression in terms of copyright. On the other hand, if creativity may not be realized because of some technical considerations, then further analysis is needed to decide if and how far copyright in the structure may still arise. Here, in the absence of relevant EU case law, the US abstraction-filtration-comparison test and criteria looked at above, may serve as a helpful pointer.

The test's application in this context may be illustrated by again referring to the case of a model of tumor growth described earlier, where a cellular bio-model is coupled to biomechanical simulations[195] As discussed this comprises a hyper-model consisting of two component models: a microscopic cellular bio-model, which shows the concentration of cells in the invaded tissue, and a macroscopic biomechanical model, which provides directions of least pressure in the tissue. This model of tumor growth is structured via interplay and exchange of information between the models. The overall function of the hyper-model is to simulate the "*geometrical evolution of the tumor predicted at the cellular level*."[196]

First, according to the test, we would dissect the hyper-model into its structural components, i.e. component models. The hyper-model structure is dictated by the two component models and a relationship in which they stand to each other. Secondly, in then eliminating the (non-copyrightable) elements dictated by efficiency, external factors, and from the public domain, respectively, the following matters would be relevant. Thus regarding efficiency, the question would be If a bio-model of cell concentration and biomechanical model of tissue are the only models, and the combination of these models by exchange of output/input data is the only way, in which the geometrical evolution of the tumor predicted at the cellular level can be simulated.

If so, the modeler´s choice of component models and their combination has merged with the underlying idea and according to the merger doctrine may not be copyrighted. However, if

[195] F. Rikhtegar, et al, 'A Model of Tumor Growth Coupling a Cellular Biomodel with Biomechanical Simulations'; see section 6.3.1.2 above.
[196] ibid.

the same process could have been achieved in another way, then the hyper-model structure may qualify as copyrightable expression. Secondly, as regards external factors, the use of elements or methods, which are "indispensable or at least standard" in the computer modeling community, would be uncopyrightable. Such elements may include: computational demands (especially for models which work on higher spatial and temporal resolutions), or requirements of Digital Model Repository (an innovative platform for exchange of cancer models), or criteria of Heterogenous Multiscale Method (HMM) etc.[197]

As we saw, the the third group of elements, which may not be copyrighted, are elements taken from the public domain. Copyright may not be claimed in "*an expression which, if not standard, then commonplace in the computer software industry*."[198] So, to the extent that the hypermodel incorporated such elements, these would similarly fall outside the scope of protection. Admittedly, the above analysis is somewhat schematic, and rather hypothetical. It is aimed to provide some general guidance and outline the elements and structures, which may be protectable by copyright, and those which may not.

6.3.4 Implications of present European IP regimes for in silico research, including in CHIC

In this concluding section, we briefly summarise the key implications that arise from the foregoing analysis of patent and copyright law, for in silico cancer modeling of the form occurring in the CHIC project. Subsequently, in the overall conclusion to this report (Chapter 7) we will use this as a starting point for suggesting possible scope for legislative reform to improve the IP rights of researchers who invest considerable skilled effort in the relevant modeling activity.

First, as regards patent law, discussed under 6.3.2, it is unlikely thatr a model from CHIC would satisfy the patenting requirements. In this regard, the need for novelty may be a particular challenge, given that other models in CHIC and related VPH projects have similar underlying conceptions, disclosed in extant academic literature, and are designed and built along similar lines. Admittedly, obtaining patent protection for the innovative solutions, generated in CHIC, is not limited to the cancer models only, hence, seeking patent protection for some inventive tools or software solutions, produced in CHIC, may well be justifiable. Even so, a serious problem remains the uncertain European legal position in relation to the meaning and scope of the excluded subject matter under Articles 52 and 53 EPO. Moreover, a practical issue remains the significant investment needed in terms of securing patent registration.

Second, as to the potential for achieving copyright protection for the modeling work, as of the current state of the European copyright law, various elements in cancer modeling, which express original creative input, may in principle be protectable by copyright. Thus, where the models are encoded into computer codes, they will enjoy copyright protection as computer programs. Such copyright in the model may also extend to the preliminary modeling work, that underlies the model code: here the modeling documentation, which records the steps, conducted in the course o f developing a model and on the basis of which such a model can be reproduced, are copyrightable. Subsequently, the almalgamated hyper-models, if

[197] T. Deisboeck, et al, 'Multiscale Cancer Modelling'.

[198] Computer Associates International, Inc. v. Altai, Inc., 982 F.2d 693, 2d Cir.1992.

structured in an innovative and original way, may count as compilations in terms of copyright. However, this question remains somewhat uncertain as it relies on analogical reasoning from other areas of IP protection (including how compliations are dealt with under the Database Directive).

More generally, the scope of copyright protection in the Software Directive, limited to literal expression of a program, i.e., the code, poses a risk, that two programs (or also computer models), confusingly similar to the user at perception level, may appear, where one program may imitate the original program as of the structure, but implementation of the code may be different. The extension of software copyright to the structure is particular relevant in view of the time and effort, which the programmer spends in designing the program. As stated above, the time and labor, which the programmer needs to study the problem, to define the modules, which would solve the problem step by step and arrange such modules into a program, by far extends the effort, which the programmer invests in coding, i.e., writing the code itself.[199]

Here EU copyright protection of software would be enhanced if protection of a computer program's non-literal expression, such as original program structure, were specifically provided for, along with the program code. Such legislative treatment could also allow for a considered policy analysis, to distinguish aspects of programming (and modeling) work that are appropriate objects for protection and reward, from elements that should remain outside the scope of protection, also taking account of relevant approaches from other jurisdictions, including the US abstraction-filtration-comparison test.

[199] WIPO expert group on the legal protection of computer software, supra; Whelan Associates, 797 F.2d 1222, 230 USPQ 481, supra.

7 Conclusion and recommendations for reforms

In this Deliverable we have considered the ethical and legal challenges posed for patient and researcher rights in relation to E-health related medical research. As has been discussed, the European regulatory framework in respect of health research and innovation has recently been subject to significant reform. This has taken the form of new or pending EU Regulations that will govern the areas of clinical drug trials, data protection law, and medical devices certification. At the same time, and notwithstanding parallel EU-level initiatives in the field of health ICT, and the considerable support offered in practice to innovative E-health projects including CHIC, there are various respects in which the research and validation environment for the new technologies still appear to be uncertain and/or suboptimal.

It is clear that with the new and pending regulatory instruments in the areas of data protection and medical devices, there will be a required period of 'bedding down', while the rules are digested and interpreted by relevant stakeholders and the courts. In this regard, it will remain to be seen if the new frameworks in the long term offer clearer and more appropriate solutions in practice for the E-health research community (while appropriately safeguarding subject/patient interests) than the previous or existing rules. However, our provisional conclusion would be that there is cause for optimism concerning data protection, but some potential cause for concern regarding medical devices.

First, as regards data protection law, we have seen that the new Regulation, when it enters force (as anticipated) in 2018, will offer a comprehensive framework that includes relevant ex ante evaluation of high risk data processing operations (such as E-health data research may generally be considered to be) inter alia through the use of DPIAs. Here the rules anticipate consultation and cooperation between researchers (be they academic or commercial) and the designated data protection officer and supervisory authorities in arriving at sensible approaches in the concrete case for using the data to its best research potential, while simultaneously taking all due care of relevant patient/data subject interests. There is in this context also encouragement for the research community to formulate autonomous codes of good processing conduct, which similarly may offer concrete guidance to researchers: moreover, and of key significance for transnational research initiatives (including much E-health-related research) a mechanism is forseen for such codes to be endorsed by the new EDPB.

We recommend that use is made of this opportunity by E-health researchers in specific fields, including in the in silico/VPH community; well-drafted codes should conduce to clear, consistent and well-thought out rules that prove easy for researchers to understand and apply on the ground. The community should seek EDPB endorsement of the code so in silico researchers across Europe may follow it in the reasonable certainty that they are thereby conforming to good ethics and the law. As noted, within the CHIC project, LUH aims to promote discussions in the direction of drafting a best practice code for in silico modelers, which at the same time utilises clustering/networking opportunities with related projects. As against this optimism, there remains, though, the risk Member States may continue in to provide in domestic law for specific safeguards in respect of health data processing. This possibility has been left to them in the Regulation (reflecting limited EU competence in this area), and only time will tell how far this may perpetuate the current, fragmented practices under the Directive.

Secondly, in relation to medical devices rules, the current draft Regulation that will at some point replace existing Directives (unlike the new Regulation on Data Protection, the precise timescale is unclear), the main risk is that a generic approach to testing and validating new clinical devices, as presently contemplated, may not be appropriate in view of specific features of in silico-based decision-support models. In particular, an approach based too strictly on the validation used for medicinal products, requiring systematic data collection from clinical trials, may prove difficult to realize in practice, and unnecessary to protect patients, as well as delaying the deployment of possible life-saving tools. We recommend that the ongoing genesis of the rules is kept under review by the Commission, and the VPH community, with a view to ensuring that the rules eventually adopted are sufficiently flexible to allow for the smooth testing and adoption of in silico predictive models, while still assuring the safety of patients.

Lastly, as to the IP regime for rewarding researchers for E-health related investment, including their creative efforts in developing and amalgamating in silico models ; this unlike the areas discussed above has not been the subject of recent/ongoing reform in Europe. However, as highlighted in this report, there are several areas where the existing rules, and their application to innovative E-health related research, must be regarded as uncertain and/or unsatisfactory. Here we would recommend legislative attention by the European law maker in relation to various matters: first, consideration should be given to a well-crafted sui generis right for researchers and others who expend time and skill in curating data to be rewarded, subject to them following stipulated mechanisms for safeguarding data privacy.

Further, in respect of patent and copyright law, attention should be directed to the best means for resolving current uncertainties in practice, by amending relevant aspects of the EPO and SWD, respectively. As regards the EPO, the policy basis for the present excludable subject matter – which may well deny protection to in silico clinical support models – should be further investigated, with a view to adopting a clear position and transparent rules on the subject. As to copyright, we recommend the creative investment in structuring of computer program, as occurs with in silico models, is recognized as a form of protectable non-literal expression within the terms of the SWD. This should clarify that the almalgamated hyper-models, if structured in an innovative and original way, qualify as compilations that also gain protection in their own right.

8 References

Alheit, K (2001): The applicability of the EU Product Liability Directive to software, The Comparative and International Law Journal, vol. 34, no. 2, 188-209

Benkler, Y (2000): 'From Consumers to Users: Shifting the Deeper Structures of Regulation', Federal Communications Law Journal 52, 561-579

Beyleveld, D, et al (eds) (2004) The Data Protection Directive and Medical Research Across Europe, Dartmouth: Ashgate.

Cheng, Y, et al (2013): Preserving user privacy from third-party applications in online social networks, In Proceedings of the 22nd International Conference on World Wide Web (WWW '13 Companion), 723-728

Cohen, I, et al (2014): 'The legal and ethical concerns that arise from using complex predictive analytics in health care', Health affairs 33, NO. 7: 1139–1147

Conrad, I, Schultze-Melling, J (2011): Beck'sches Mandats Handbuch IT-Recht, München: C.H. Beck Verlag

COM eHealth Task Force (2012) 'Redesigning health in Europe for 2020'

Coveney, P, et al. (2014) Computational Biomedicine, Oxford University Press.

Deisboeck, Z. et al (2011): 'Multiscale Cancer Modelling', Annual Review of Biomedical Engineering, vol. 13, pp. 127-155, August 2011, doi: 10.1146/annurev-bioeng-071910-124729

Dreier, T (2009): 'Regulating information: Some thoughts on a perhaps not quite so new way of looking at intellectual property', in: J. Drexl et al (eds), Technology and competition-technologie et concurrence, Bruxelles

European Commission (2014): Green Paper on mobile health ('mHealth'), COM(2014)

Forgó, N, et al., (2010): Ethical and Legal Requirements for Transnational Genetic Research, München: C.H.Beck Verlag

Forgó, N (2015): 'My health data—your research: some preliminary thoughts on different values in the General Data Protection Regulation', International Data Privacy Law 5 (1) 54-63.

Herveg, J, et al (2007): Final Recommendations on Legal Issues in Ehealth (Legally eHealth project, D.5) COM 30-CE-004 1734/00-55

Hoeren, T (2014): IT-Recht, München: C.H. Beck

Kizza, J (2010): Ethical and social issues in the information age, London: Springer

Lessig, L (2002): The Future of Ideas, The Fate of the Commons in a Connected World, New York

Lloyd, I (2008): Information Technology Law, Oxford: Oxford University Press

Lowrance, W (2002) Learning from Experience: Privacy and the Secondary Use of Data in Health Research, London: the Nuffield Trust

Marly, J (2014): Praxishandbuch Softwarerecht – Rechtsschutz und Vertragsgestaltung, München: C.H. Beck

Pedroni, J and Pimple, K (2001): A Brief Introduction to Informed Consent in Research with Human Subjects, Poynter Center, University of Indiana

Peterson, I, et al (2015): Systems Biology, Information Technology and Cancer Research in M. Döring et al., Contextualizing Systems Biology, Springer, Switzerland

Peukert, A (2008): ‚Güterzuordnung und Freiheitsschutz', in: J. Hilty et al (eds), Geistiges Eigentum, Herausforderung Durchsetzung, Berlin

Rikhtegar, F, et al (2014): 'A Model of Tumor Growth Coupling a Cellular Biomodel with Biomechanical Simulations", In G. Stamatakos and D. Dionysiou (eds): Proc. 6th Int. Adv. Res. Workshop on In Silico Oncology and Cancer Investigation – The CHIC Project Workshop (IARWISOCI)

Rowland, D,. et al, (2012): Information Technology Law, Abingdon: Routledge

Sanchez, AN (2016): 'Privacy by default' and active 'informed consent' by layers: essential measures to protect ICT users' privacy', J. of Information, Communication and Ethics in Society 14 (2)

Stamatakos, G, et al. (2014) 'The technologically integrated oncosimulator: combining multiscale cancer modeling with information technology in the in silico oncology context', J. Biomed Health Inform. 18(3):840-54

Stenzhorn, H, et al (2010): 'The ObTiMA system - ontology-based managing of clinical trials', Stud Health Technol Inform. 2010; 160(Pt 2):1090-4

STEP Consortium, Seeding the EuroPhysiome: A Roadmap to the Virtual Physiological Human

Stokes, S (2014): Digital Copyright, Law and Practice, 4th Edition, Hart Publishing, Oxford and Portland, Oregon

Tartarini, D, et al (2014): 'The VPH Hypermodelling Framework for Cancer Multiscale Models in the Clinical Practice', In G. Stamatakos and D. Dionysiou (Eds): Proc. 2014 6th Int. Adv. Res. Workshop on In Silico Oncology and Cancer Investigation – The CHIC Project Workshop (IARWISOCI)

Tomita, M. (2001) 'Whole-cell simulation: a grand challenge of the 21st century', Trends Biotechnol. 19(6):205-10

US Food and Drug Administration (2013) 'Paving the Way for Personalized Medicine'

Viceconti, M (2011): "A tentative taxonomy for predictive models in relation to their falsifiability", Philos. Transact A Math Phys Eng Sci, vol. 369, pp. 4149-61, 1954

Vihul, L (2014): The liability of software manufacturers for defective products, The Tallin Papers, vol. 1, 1-14

Wagner, A (1999): Binäre Information als Gegenstand des Rechtsverkehrs, Norderstedt

Weber, R (2002) ‚Ali Baba oder das Risiko exklusiver Informationsinhaltsrechte', in: R. Schweizer et al (eds), Festschrift für Jean Nicolas Druey zum 65. Geburtstag, Zürich 2002,

Zech, H (2012): Information als Schutzgegenstand, JUS PRIVATUM, Beiträge zum Privatrecht, Band 166, Mohr Siebeck Tübingen

Appendix 1 – Abbreviations and acronyms

ACGT	Advancing Clinico-Genomic Trials on Cancer
API	Application Interface
CHIC	Computational Horizons In Cancer
CJEU	Court of Justice
COM	European Commission
CTD	Clinical Trials Directive
CTP	Clinical Trial Protocol
CTR	Clinical Trials Regulation
DoH	Declaration of Helsinki
DPIA	Data Protection Impact Assessment
DPD	Data Protection Directive
DoW	Description of Work
EC	European Community
ECHR	European Convention on Human Rights
EDPB	European Data Protection Board
EHR	Electronic Health Record
EPC	European Patent Convention
ERN	European Reference Network
EU	European Union
FP7	Seventh Framework Programme
GA	Grant Agreement
GDPR	General Data Protection Regulation
HIS	Hospital Information System
HTTP	The Hypertext Transfer Protocol
HTTPS	HTTP over Secure Socket Layer
ICT	Information and Communication Technology

IP	Intellectual Property
IPR	Intellectual Property Right
ICCS	Institute of Communications and Computer Systems
LEC	Local Ethics Committee
LUH	Leibniz Universität Hannover
MDD	Medical Devices Directive
MDR	Medical Devices Regulation
R&D	Research and Development
RCT	Randomised Control Trial
SWD	Software Directive
TRIPS	Trade-Related Aspects of Intellectual Property Rights
UK	United Kingdom
VPH	Virtual Physiological Human
WIPO	World Intellectual Property Organization
WMA	World Medical Association