



Deliverable No. 4.3.2
Development of the data protection
and copyright framework for CHIC
second iteration

Grant Agreement No.: 600841
 Deliverable No.: D4.3.2
 Deliverable Name: Development of the data protection and copyright framework for CHIC second iteration
 Contractual Submission Date: 30/09/2016
 Actual Submission Date: 30/09/2016

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.:	D4.3.2
Document name:	Development of the data protection and copyright framework for CHIC second iteration
Nature (R, P, D, O) ¹	R
Dissemination Level (PU, PP, RE, CO) ²	PU
Version:	1.0
Actual Submission Date:	30/09/2016
Editor: Institution: E-Mail:	Nikolaus Forgó LUH forgo@iri.uni-hannover.de

ABSTRACT:

This deliverable is the second of a two part iteration deliverable (following up D4.3.1) that describes the data protection, data security and intellectual property rights framework developed for the CHIC project. It also builds on the other previous legal ethical deliverables – D4.1, D4.2, and D4.4, in showing the concrete legal, organisational and technical measures put in place to safeguard the medical data used for the project, now with particular reference to the CHIC project validation and exploitation phases. The measures include patient consent, data protection agreements to be concluded between project partners and other validating users, and a dedicated data security framework protecting the data repositories and flows during validation. Also considered is the required certification of in silico models under the medical devices regime.

Regarding copyright issues, an analysis is made of the protection potentially available in the project, which takes account of the specific contributions of relevant parties. Also, to support the developing parties with making their license choices and to mitigate the potential license incompatibility risks, the software components and models in CHIC are analysed on the subject of license incompatibility issues and the results presented in a software licensing report

KEYWORD LIST:

Data Protection, Data Security, Copyright; Intellectual Property Rights

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

The author is solely responsible for its content, it does not represent the opinion of the European Community and the Community is not responsible for any use that might be made of data appearing therein.

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)

MODIFICATION CONTROL			
Version	Date	Status	Author
0.1	27/08/2016	Draft	Marc Stauch, LUH
0.2	06/09/2016	Draft	Iryna Lishchuk, Marc Stauch, LUH
0.3	12/09/2016	Draft	Elias Neri, Custodix, Iryna Lishchuk, LUH,
0.4	16/09/2016	Draft	Marc Stauch, LUH, Elias Neri, Custodix
0.5	20/09/2016	Draft	Nikolaus Forgò, LUH
0.6	22/09/2016	Draft	Marc Stauch LUH, Elias Neri, Custodix
0.7	27/09/2016	Internal Review	Norbert Graf, USAAR, Georgios Stamatakos, Nikolaos Tousert, ICCS, Debora Testi, CINECA
1.0	30/09/2016	Final	Marc Stauch, Nikolaus Forgò, LUH

List of contributors

- Elias Neri, Custodix
- Norbert Graf, USAAR
- Georgios Stamatakos, ICCS
- Nikolaos Tousert, ICCS
- Debora Testi, CINECA
- Iryna Lishchuk, LUH
- Marc Stauch, LUH
- Nikolaus Forgò, LUH

Contents

1	EXECUTIVE SUMMARY	5
2	INTRODUCTION	6
3	STRUCTURE	7
4.	CHIC DATA PROTECTION FRAMEWORK – SECOND ITERATION	8
4.1	BACKGROUND	8
4.2	FURTHER CONSIDERATIONS ARISING DURING CHIC MODEL VALIDATION PHASE	9
4.3	PATIENT/PROXY CONSENT	11
4.4	CONTRACTUAL FRAMEWORK BETWEEN CLINICIANS, CHIC INFRASTRUCTURE, AND MODELERS	12
4.4.1	CHIC Clinical User Agreements	13
4.4.2	CHIC End User Agreements	14
4.4.3	CHIC Data Security Agreement	15
4.4.4	Further regulatory requirements for validation of models	16
5.	CHIC DATA DE-IDENTIFICATION PROCESS	19
5.1	INTRODUCTION	19
5.2	DE-IDENTIFICATION OF DATA SETS – A CONTEXT RELATED UNDERTAKING	19
5.3	ROLE OF STANDARDS	20
5.4.	DE-IDENTIFICATION APPROACH IN CHIC	21
5.4.1	CHIC data de-identification flow	21
5.4.2	Using Syntactic De-Identification Techniques	21
5.4.3	Overview of CHIC de-identified datasets	23
5.4.4	CHIC Data de-identification process	25
5.5	CUSTODIX ANONYMISATION TOOL SERVICES (CATS)	30
5.5.1	Custodix PseudoEngine	30
5.5.2	Custodix Privacy Language	34
6.	INTELLECTUAL PROPERTY ASPECTS	39
6.1	CHIC EXPLOITABLE OUTPUTS UND SOFTWARE LICENSING	39
6.1.1	CHIC exploitable outcomes	39
6.1.2	Software licensing issues	40
6.1.3	Licensing solutions for CHIC software components	40
6.1.4	Licensing solutions for CHIC models	41
6.1.5	Licensing solutions for CHIC hyper-models	43
6.1.6	Licensing solutions for the CHIC repository architectures	43
6.1.7	Suggested open source licenses	44
6.1.8	Legal nature and licensing solution for the CHIC platform	46
6.2	MANAGING IPR IN CHIC	47
6.2.1	CHIC Memorandum of Understanding	47
6.2.2	Managing IPR in hyper-models	47
6.2.3	Access Rights to hyper-models	48
6.3	PROTECTION OF CHIC REPOSITORIES BY SUI GENERIS DATABASE RIGHTS	49
6.3.1	CHIC repositories	49
6.3.2	Sui generis database rights	49
6.3.3	Applicability of sui generis database rights to the CHIC repositories	52
6.3.4	Sui generis database rights and rights in software and data	55
6.4	PROTECTABILITY OF CLINICAL DATA BY IP RIGHTS	56
6.4.1	Undisclosed information	57
6.4.2	Contractual approaches	58
6.4.3	Data ownership issues	58
7.	CONCLUSION	60
8	REFERENCES	61
	Appendix 1 – Abbreviations and acronyms	63
	Appendix 2 – Patient Consent to Data Use for Model Validation	64
	Appendix 3 – CHIC Clinical User Agreement	68
	Appendix 4 – CHIC End User Agreement	76
	Appendix 5 – CHIC Data Security Agreement	81
	Appendix 6 – CHIC software licensing report	85

1 Executive Summary

The CHIC project aims at developing cutting edge ICT tools, services and secure infrastructure to foster the development of elaborate and reusable integrative models (hypermodels) in the field of cancer diagnosis and treatment, as well as larger repositories so as to demonstrate benefits of having both the multiscale data and the corresponding models readily available in the VPH domain. In the course of developing these tools, both retrospective and prospective patient data are being used to test these models as well as validate them. In Deliverable D4.3.1 (submitted at PM14 in May 2014), the legal and ethical framework for the processing of this data was implemented in the form of a set of contracts, establishing a closed 'network of trust' between the project partners, backed up by secure de-identification and technical safeguards for data entering and stored within the project infrastructure, administered by the project's data security partner. The present deliverable, which provides the second iteration of that framework, adds to it by supplying the legal and ethical basis to allow clinicians (including potentially ones not previously involved in the project) to begin early validation in the context of clinical care of the usability of the models and utility of the predictions they provide, subject to the fully informed consent of the patients concerned and the obtaining of appropriate ethics committee approval. In this regard it also takes account of the impending changes in the European data protection framework signalled by the adoption of the General Data Protection Regulation 2016/679 (due to enter force in May 2018).

In the second place, this deliverable provides a follow-up analysis, to that of Deliverable D4.3.1, of the copyright framework relating to the development of the CHIC models and tools. Here we examine relevant copyright and IPR aspects that have particular implications for the transition of the project into the exploitation stage. Exploitable software outcomes of CHIC are identified and suitable licensing options are suggested, both for the CHIC components, repository architectures, models and hyper-models, and for CHIC integrative platform as a whole. Further, we analyse the protectability of the CHIC repositories by sui generis database rights and the scope of such protection, as well as alternative options of protecting the clinical data in CHIC by IP rights and contractual mechanisms. To address potential license incompatibility issues, software component and model dependencies are analysed and presented in the CHIC software licensing report

2 Introduction

This document describes the data protection, data security and IPR framework developed for the CHIC project. It is the second of a two part iteration deliverable (the first being deliverable D4.3.1) that outlines the concrete data protection and copyright framework. It also builds on the other previous legal ethical deliverables – D4.1, D4.2, and D4.4, in showing the concrete legal, organisational and technical measures put in place to safeguard the medical data used for the project, now with particular reference to the CHIC project validation and exploitation phases. Crucially, at this second iteration stage, where clinicians will deploy the models (and compare the outcomes they predict with real patient progress) the relevant clinicians will be processing the identifiable personal data of their patients. This contrasts with the use of clinical data in the model-building phase during the project, in which data was securely de-identified prior to its use in the course of developing the models to begin with.

Accordingly it will be essential, from both a legal and ethical point of view, for the clinician to obtain the patient's fully informed and explicit consent to the use of the patient's data for this purpose; here the deliverable includes a draft patient information sheet and consent for the clinician to use. In addition, provision is made for the new data and information generated by the validation process itself to be fed back (in securely de-identified form) to the modelling partners to use in further developing and refining the models. This aspect will be addressed by continued reliance on binding data protection agreements between the relevant project partners, backed up by the data security framework to protect the data repositories and flows during validation.

This document further describes the data de-identification process as performed over the life time of the project. This includes an overview of all datasets de-identified within the project's scope and uploaded to the CHIC clinical data repository, as well as an abstract description of how those data sets have been de-identified. The report does not go into granular details of the de-identification processes as this would be an infringement of the data protection framework as defined by the project and described in D.4.3.1.

Regarding IPR and copyright issues, the present analysis builds upon that offered in Deliverable D4.3.1, and presents guidelines for the transition of the project into the exploitation stage. This includes identification of exploitable software outcomes of CHIC and analysis of licensing options, for the CHIC components, repository architectures, models and hyper-models, and the CHIC integrative platform as a whole. To support the developing parties with making their license choices and to mitigate the potential license incompatibility risks, the software components and models in CHIC are analysed in relation to potential license incompatibility issues and the results presented in the software licensing report

3 Structure

The deliverable describes the second iteration of the data protection and copyright framework of the CHIC project. It is divided into two broad parts. The first part consists of chapters 4 and 5, and presents the data protection de-identification framework, as revised so as to accommodate the clinical validation of the models by clinicians using personal patient data. Here the legal and organisational measures required to allow for use of prospective patient data by the clinicians and subsequently (in securely de-identified form) by the modellers as part of project exploitation are detailed. Compared to the model design phase, covered in the first iteration, which relied upon retrospective, de-identified clinical data, the clinician will make use of the patient data in the context of a direct clinical care relationship, where the need for clear understanding by the relevant patients of what this entails for their treatment is legally and ethically imperative.

Accordingly, besides the ongoing need for data protection agreements to safeguard the transfer of data to the CHIC platform infrastructure, a model patient information sheet and consent form to be used by the clinicians has been drafted and is annexed to the deliverable in Appendix 2. In this regard, it will be important to distinguish interactions between the clinician and the patient that would fall under the rubric of ordinary clinical care and those where the basis for a given action by the clinician (including the use and/or transfer of data collected from the patient) lies in the project exploitation and research. The technical de-identification measures that have been equally essential for the proper and effective functioning of the data protection framework (and will continue to be so during model validation and exploitation) are then elaborated in Chapter 5.

The second main part of the deliverable, chapter 6, describes the copyright framework of the project. This expands upon the analysis presented in Deliverable D4.3.1, by presenting guidelines for transition of the project into the exploitation stage. This includes the identification of exploitable software outcomes of CHIC and analysis of suitable licensing options, for the CHIC components, repository architectures, models and hyper-models, as well as for CHIC integrative platform as a whole.

At the end of the document, in Appendices 2-6, the model patient information sheet and consent and the draft second iteration data protection agreements are then presented, together with the CHIC software licensing report.

4. CHIC Data Protection Framework – Second Iteration

4.1 Background

As set out in Deliverable D4.3.1, the CHIC first iteration framework was designed to ensure the secure, legal and ethical use of retrospective patient data for the purpose of building and developing the models and hypermodels in accordance with the project aims. Following this approach, the project has made use of retrospective patient data on the four different cancers under study (Wilms tumour, Glioblastoma multiforme, prostate cancer, and non small cell lung cancer, which have been undergoing processing within the CHIC environment. The data in question has included clinical data, imaging data, molecular data, metadata, annotations, added semantic information to data and model / hypermodel configuration parameters.³

An important feature of such data use, which served as a starting point for the data protection framework, was that the modellers did not require data in identifiable form in order to develop the models; rather the models seek to generalise from multiple patient datasets by associating anonymous data values and patient outcomes in order to identify the algorithms informing future model interactions and predictions. In line with the principle of ‘data minimisation’ contained in the Data Protection Directive 95/46/EC (and reaffirmed in the new General Data Protection Regulation 2016/679⁴), according to which data should be processed so far as is possible – compatibly with the processing purpose – in non-personal form, the first iteration framework provided for rigorous and secure de-identification of all data. Here the CHIC project has made use of a tried and trusted approach, which had previously been successfully deployed in a number of other EU projects, where it was also important to work with real clinical data after setting up the platform, namely ACGT, P-Medicine, and EURECA.

The basic assumption of the relevant data protection framework is that the best way to safeguard patients’ rights would be achieved, if only anonymous data were processed in the project. At the same time, as discussed in Deliverable D4.3.1, it was recognised that – even after the removal of the more obvious patient identifiers – the absolute anonymity of clinical patient level data (as opposed to purely statistical data) cannot be guaranteed due to the fact that rare (and potentially individuating) values may need to be retained in the datasets to meet the needs of the project.

The framework in question incorporates secure technical, organisational and legal measures to so far as possible eradicate risks to the privacy and/or autonomy interests of relevant patient subjects, and which consists in three core ‘pillars’. In the first place a set of contracts have been concluded between the project partners and the Center for Data Protection (CDP) a legal entity representing the consortium, providing amongst others data protection policies, clauses on liability, in case data is unlawfully matched or disclosed, as well as provisions to ensure the safe disposal of data once it is no longer required for the purposes of the Project. Secondly, this has been backed up by a security infrastructure, including dedicated de-identification software (CAT) and user-identification and authentication services. These measures together aim to ensure that re-identification of the retrospective patient data in the project is not possible with means likely reasonably to be used with respect to time, expense and labour. Thirdly, and in order to ensure the project proceeds in full compliance with external data protection and ethical requirements, the clinical partners providing data to the project (USAAR, UNITO, KU Leuven) obligated themselves, as a term of the agreements under which they provided the data, to ensure they had obtained the relevant ethics

³ See the Description of Work (DOW) Part B, pp. 5-8.

⁴ Official Journal L119/1; enacted 27 April 2016 and due to enter force on 25 May 2018; the text of the Regulation is available at: [http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf].

committee approval to do so. This was important as the use of the data for CHIC involves further processing purposes to those (treatment and/or different specified research projects) for which the patient originally consented to the collection of their data.

As noted, the above agreements were designed to take care of the model development phase of the project, and will operate within a closed community of researchers. At the present stage (M42) of the project, though, and as foreseen in the description of work, a new second iteration framework is needed. In particular, this must take account of the fact that, in order to start to validate the models and hypermodels being developed in the project, and for their exploitation, clinicians will process the identifiable personal data of their patients when they deploy the models (and compare the outcomes they predict with real patient progress).

An additional aspect is that, whereas previously the processing of data in the project involved a closed set of clinical users (from project partner institutions) it is desirable, during the validation and exploitation phase, for the infrastructure to be open to external clinicians, who would want to get some treatment predictions based on the developed models and hypermodels. Accordingly, new service level agreements are required for such external users, who would upload data into the CHIC infrastructure, and be responsible for obtaining the consent of the data subjects or any other approval that may be required in that case. For its part the CHIC infrastructure will maintain the technical, organisational and security features that have been embedded into the design of the system to protect data. A final point is that, as noted above, the existing data protection legislation (EU Directive 95/46/EC) is due to be replaced in May 2018 by the General Data Protection Regulation 2016/679 ('GDPR'), raising the question (previously touched on in Deliverable D4.4⁵) of how the new provisions may affect the ongoing future use of patient data in validating *in silico* models.

4.2 Further considerations arising during CHIC model validation phase

The projected manner in which the CHIC models and hypermodels will be deployed by clinicians for the purposes of user-testing and validation has been detailed in Deliverable D2.5. As noted there, a major concern of the project at this stage is to demonstrate the clinical relevance of the developed hypermodels and how they will and can be used beyond the lifetime of the project. Within the CRAF platform environment described in the document, the clinicians will act as drivers in an iterative process – featuring feedback loops between them and the modellers – to ensure the resulting models are adequately user-friendly (from the perspective of an average experienced clinical oncologist) and also that they show *prima facie* ability to provide clinicians with information for optimising treatment of their patients (as summarised in the clinical questions per cancer-type in deliverable D2.5), so as to offer an advance over existing standard treatment approaches.

As regards the use of patient data, there will, as already mentioned, be an important difference at this stage compared to the earlier use of data when building the models. At that point, the modeller partners were using retrospective de-identified/anonymous data of patients to construct the algorithms to be used by the models. Some preliminary validation of the models will be done by the modelers, still using anonymous and retrospective data. This is in accord with standard data analytical procedure (involving the use of hold-out datasets, etc). However, for further validation purposes, and at the point at which clinicians in CHIC are involved, individual prospective patient data will be used: here the target value of interest: e.g. tumour shrinkage post-chemotherapy) is as yet unknown; the question is how well the models (based on algorithms generated by mapping associations between data at diagnosis and treatment outcome in the retrospective patient datasets) can predict what will likely happen in new prospective patients (for whom *ex hypothesi* only the data at diagnosis is available).

⁵ CHIC Deliverable D4.4 (March 2016), part 4.4.

For these purposes, it is envisaged that the clinician will interact with newly diagnosed patients and process their data in accordance with the following steps:

- (i) Clinician with new patient uploads data at diagnosis (for nephroblastoma, GBM, NSCLC, the relevant data goes to ObTiMA; in case of prostate cancer, to the Eureka1/2 database) for curation/pre-processing;
- (ii) Subsequently the patient data is transferred via the CHIC clinical research application framework (CRAF) to the CHIC clinical data repository;
- (iii) Clinician then selects a suitable model within the CRAF to use for that patient (depending on the patient cancer-type and available data); on pressing the command 'run', the system executes the model populated with the patient data to generate a prediction (e.g. "chemotherapy will achieve 20% tumor reduction");
- (iv) The model prediction is placed (together with metadata relating to the execution of the model) in the CHIC in silico trial repository, where the clinician is able to view it.

Subsequently (v), the clinician may, once the real patient outcome is known, transfer this information together with the outcome predicted by the model to the CHIC research domain; the modelers would be able to access this data (in securely de-identified form) in order to check how the model has performed (how far does the predicted outcome accord with the real outcome?) and adapt and make refinements to the model based on this.

The flow of patient data that occurs between the clinician, the CHIC infrastructure, and the modelers is represented schematically in Figure 1 below:

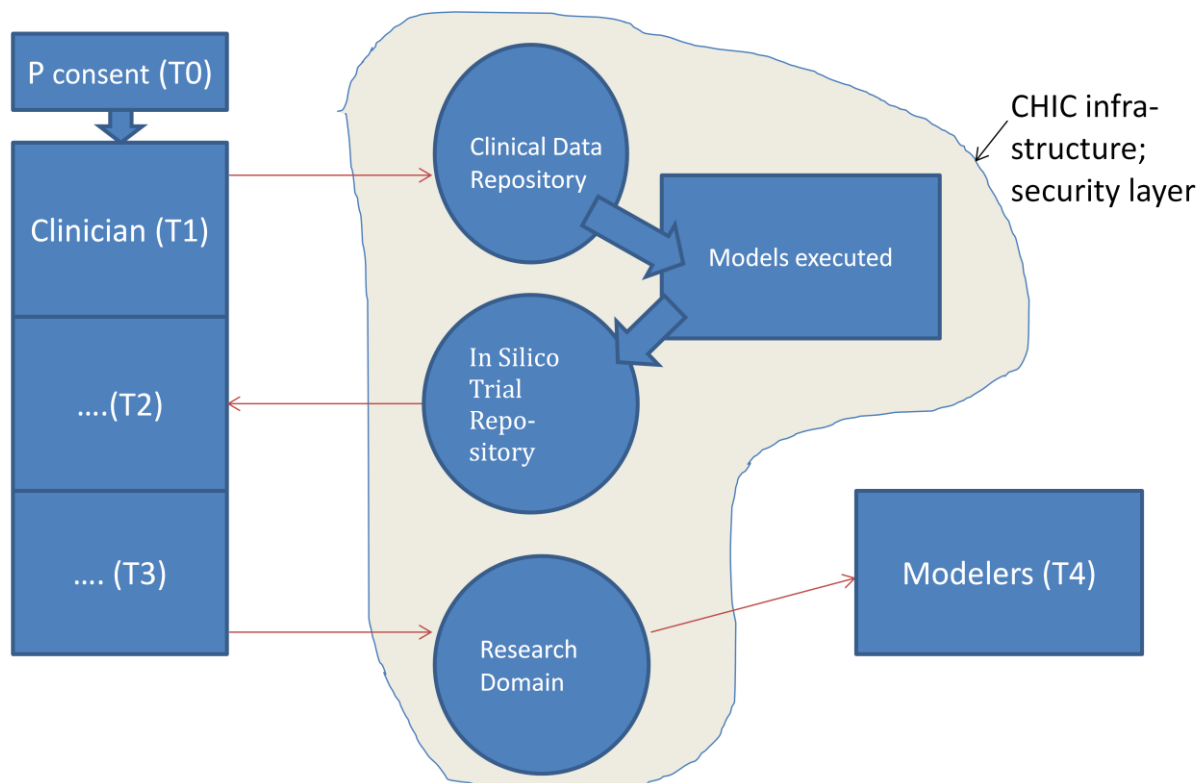


Figure 1: Data Use in CHIC Model Validation Phase

Here, in terms of characterizing the data-flows, the initial transfer of data from the clinician to the CHIC infrastructure (steps (i) and (ii)) make up T1; step (iii) comprises an internal data processing operation within the CHIC infrastructure, but step (iv), when the clinician

accesses new information about the patient in the in silico trial repository, viz. the patient outcome predicted by the model, represents in effect a data transfer back from the CHIC infrastructure to the clinician (T2); later, where the clinician sends the predicted and real outcomes to the research domain, this amounts to a further transfer (T3) to the CHIC infrastructure; finally, when the modelers access the latter data to use for refining the model, this constitutes a data transfer (T4) from the infrastructure to the modelers.

The legal conditions for implementing this validation process will be addressed in the ensuing subparts of this Chapter. However, it is pertinent to mention a number of further points here to clarify and anticipate that discussion. First, as shown in the diagram, at T0, prior to transferring the data, it is essential that the clinician should have obtained the patient's consent to the data processing in question. This aspect, and the matters on which the patient should be informed when providing such consent, are looked at in part 4.3.1 below. Second, and in line with the legal principle of data minimization discussed in part 4.1 above, the data needs to be in a different form when processed by the clinician in the course of treating the patient (T1 and T2), as opposed to when it is later uploaded by the clinician to the research domain for use by the modelers (T3 and T4).

More specifically, the data will be single-pseudonymised when uploaded and stored in clinical data repository, where the clinician is able to access it; following execution of the model, the prediction for the patient will be placed at T2 in the in silico trial repository, in the same single-pseudonymised form ("Prediction for Patient XW567: 20% tumour reduction"), for the clinician to view. It is important to emphasize that only the clinician will have access to this data (and for his own patients only), so as to be able to compare what the model predicts concerning the particular patient (in his care) with his observations as to what actually occurs. Moreover, in the hands of the clinician, the data is protected by the latter's strong fiduciary duty of confidentiality. By contrast, at T3, when the clinician transfers the data, plus the actual patient outcome, to the research domain for the modelers to access and use (T4), the data will (when passing through the CHIC security infrastructure) be subject to double-pseudonymisation plus other de-identification measures in the same way as the retrospective data used earlier for building the models. This reflects the fact that (unlike the treating clinician) there is no justification for the modelers to be able to link the data to any individual patient; they need simply to know how far the model, run in a given manner (recorded in the metadata), generated a predicted outcome divergent from the actual outcome.

Two final, inter-linked issues relate to the manner in which the clinicians at T2 will be using the model, and how far this may be subject to other (non-data protection related) regulatory oversight. First, at this stage, it is clearly not envisaged that a clinician, who receives a model prediction for his patient, would rely on this when deciding what treatment to actually give the patient. It is evident that to do so, at a point when the accuracy of the models has ex hypothesi yet to be tested, would be unethical risk-taking and indeed amount to clinical malpractice. Rather, the clinician's role will simply be to use the model (to ascertain user-friendliness) and observe and feed back the model's success-rate to the CHIC infrastructure, as described at T3. Nonetheless, given the possible risk of unconscious reliance, it is in our view desirable that the clinician should (in addition to patient consent) also obtain ethics committee approval. Secondly, a question arises as to the applicability already at this point of the medical device regime for certifying medical products (including stand-alone software systems) for marketability. Both of these issues are considered further under 4.3.3 below.

4.3 Patient/Proxy Consent

The expectation is that in every case the clinician and/or hospital institution will obtain the patient's (or his or her legal proxy's) informed consent to take part in the validation of the CHIC models and hypermodels. Here, the assumption is that such participation will not involve any additional physical procedures or interactions (such as extra collection of data or

samples), to those the patient will undergo in any event, either as part of the standard therapy or (where the patient is enrolled in a clinical trial following diagnosis of their condition) as part of that trial. It follows that the implications for the patient of agreeing (at same time) to take part in the CHIC validation relate solely to the additional processing of their data (collected in any event) required for that purpose.

Nevertheless, as discussed in detail in deliverable D4.1,⁶ what is at issue here is the processing of highly sensitive personal data, with significant potential risks for the patient's privacy and autonomy. Here both ethical guidance, deriving from the need to protect the patient's fundamental interests,⁷ and data protection law are clear that, wherever practical, the patient needs to consent to the relevant data processing. Indeed, in the case of sensitive health data processing, Directive 95/46/EC imposes a requirement for the consent to be both explicit and specific.⁸ According to the Article 29 Working Party (set up under the Directive) 'specific' should be read as entailing a detailed understanding on the patient's part of the concrete situation in which his or her data will be used.⁹ Under the forthcoming General Data Protection Regulation, it seems certain that the conditions will be similarly interpreted.¹⁰

A *Model Patient Information Sheet and Consent Form* has been drafted for this purpose, and is included in Appendix 2 of this Deliverable. In our view this document would serve to inform patients sufficiently in legal and ethical terms of the relevant implications (including possible risks) of such a step to give valid informed consent to the processing of their data for validating the CHIC models. Importantly, the consent form distinguishes between two main stages at which data is processed, as these are in principle distinct: first, as discussed in part 4.2 above, the data is processed principally by the clinician, when transferring the data to CHIC clinical data repository, executing a model, and accessing the predicted outcome information in the CHIC in silico trial repository (T1 and T2); subsequently, as we saw, the plan is for the clinician to feed this data, now in securely de-identified form, into the CHIC research domain for the modelers to utilize (T3 and T4).

Nonetheless, it is possible that a patient, while happy for his or her clinician to see and use the data, may not be comfortable with the further use of the data by the modelers for extended purposes. Accordingly, the patient is given the option to consent to (or decline) the second main stage of processing. A further point is that in some cases, notably those where the data of child patients is used for clinically validating the nephroblastoma models, the patient will be legally incapable of giving consent. In such a case, consent may (and must) be provided by the patient's legal guardian (usually a parent) acting as a proxy,¹¹ and the consent form has accordingly been designed to allow for this.

4.4 Contractual framework between clinicians, CHIC infrastructure, and modelers

As explained in part 4.1 above, the legal and organisation component of the CHIC data protection framework in its first iteration has been founded on three mutually supporting contractual agreements: the CHIC Data Provider Agreement, signed by the clinical project partners providing patient data to the CHIC infrastructure; the CHIC End User Agreement, signed by the technical modelling partners, accessing and processing the data within CHIC; and the CHIC Trusted Third Party (TTP) Agreement, signed by the data security provider

⁶ Deliverable D4.1, Initial analysis of the ethical and legal requirements for the sharing of data (September 2013), sections 5.2.2 and 6.2.

⁷ See e.g. the WMA Declaration of Helsinki (2013 revision), article 26.

⁸ Directive 95/46/EC, articles 8(2)(a) and 2(h). This will be maintained by the GDPR, articles 4(11) and 9(2)(a).

⁹ See http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2007/wp131_en.pdf, at 8; see also the Art 29 Working Party's 2011 opinion on the definition of consent (15/2011; WP187), at 25 ff.

¹⁰ The GDPR in article 4(11) adopts identical language to the Directive, when defining consent.

¹¹ See the WMA Declaration of Helsinki (2013 revision), article 28.

(the project's data security partner, Custodix) acting as trusted third party. The other party to these respective contracts is the Center for Data Protection (CDP); the latter operates as the central data controller for the project, ensuring compliance with the data protection and security framework established for the project. The utilization of the CDP in this way has (as it did in earlier related EU projects, notably ACGT and P-Medicine¹²) solved the need for a legal body in lieu of the consortium, with the capability of concluding binding data protection contracts.

As discussed in Deliverable D4.3.1,¹³ the underlying rationale for these contracts is to create a closed user community ('network of trust') in which the various parties enter into explicit reciprocal obligations to each other concerning their respective spheres of responsibility for the lawful and secure use of clinical data within the project: thus data providers take clear responsibility vis a vis the data users as to the prior conditions (such as patient consent, and initial pseudonymisation) permitting data transfer to the project, whereas the technical partners in return agree to handle the data at all times in a safe and secure way and for the specific and limited purposes countenanced by the project.

In terms of the validation of the models that will now occur, a similar contractual framework is equally indicated and will be utilised. Indeed, as described in part 4.1, some of the data flows at this stage will remain functionally equivalent to those that occurred previously when building the models. This is the case with regard to the second stage data transfers (T3 and T4) which occur between the same parties (clinicians/clinical institutions as providers, CHIC as the intermediary, modelers as users), and for very similar purposes (continuing development of the models) as the use of securely de-identified retrospective clinical data during the model building phase. Nonetheless, it is apparent that some modifications will also be required, not least to cover the transfers (T1 and T2) of identifiable (pseudonymised) patient data between clinicians and the CHIC infrastructure.

Below we describe the key changes to the respective contracts (as opposed to the provisions of the analogous first iteration contracts) for them to operate as part of the second iteration data protection framework. The existing (first iteration) contracts will remain in force until the end of the CHIC project in March 2017, after which the modified second iteration agreements would govern the further validation and exploitation process. However, the modified contract between the clinician and CHIC infrastructure has been drafted also to cover possible clinical validation of models that begins before then, assuming the required patient consent and other approvals are obtained. Particularly here there are (compared to with the earlier CHIC data provider agreement governing provision of clinical data to the project) some key changes.

4.4.1 CHIC Clinical User Agreements

As just mentioned, the most significant changes will be required to the agreement between the clinical providers of patient data and the CHIC infrastructure: this reflects that the clinicians will at this stage would provide prospective patient data, and indeed also now be using it to validate the models; moreover the data itself (inside the CHIC infrastructure) will remain accessible to them in identifiable personal form. At this point, the clinical providers would indeed retain a large measure of control over the way the data is processed (by choosing the model to run, populating the model with the patient data, and viewing the executed model outcome. At a conceptual level, this may also seem to provide an argument for regarding the CHIC infrastructure as simply a 'data processor' for the purposes of EU data protection legislation,¹⁴ acting pursuant to the choices and purposes of the clinicians (as

¹² See [acgt.ercim.eu/]; and [www.p-medicine.eu/].

¹³ Deliverable D4.3.1, section 4.3.

¹⁴ Under Directive 95/46/EC, Article 2 (e).

data controllers). Thus, the defining elements in the controller-processor relationship are that the controller assumes overall responsibility for the processing occurring, and also that it determines the purpose of the processing.¹⁵ The better view, though, is that both parties act as joint data controllers with regard to the operations performed on the data. This reflects their underlying common purpose to work towards model validation. Indeed, as a further aspect of this, the CRAF platform will also collect metadata (which the clinician himself will not see) that records the way the clinician deployed a given model; the reason for this is to provide information for later – in strictly de-identified form – by the modelers when refining the models.

In the light of the above, besides changes to the preamble to describe the data processing planned at this stage in validating the models, a number of new clauses are included. These, inter alia, detail further obligations on the clinical partners (in subclauses of Clause 3) in terms of having obtained patient consent and other required regulatory approvals¹⁶ to carry out the validation, as well as the duty to exercise care and implement safeguards when processing the data. Changes are also made in Clause 4 to the obligations of the CDP (or future administrator of the CHIC infrastructure), in particular to implement the technical and organisational measures necessary to secure a strict separation between the patient data in the clinical data and in silico trial repositories (accessed and used by the clinicians in single-pseudonymised form) and the securely de-identified data in the research domain (accessed and used by the modelers). In this regard the personal data should be deleted from the CHIC infrastructure as soon as the clinician has finished using it.

A further issue concerns the metadata generated by the CHIC system for use by the modellers. Here, in view of the separate patient consent required to allow processing of the de-identified patient data (discussed in part 4.3.1 above), a requirement is imposed on the CHIC administrator to check that such consent has been provided before making the associated metadata available to the modelers. A final point concerns the potential parties to the agreement. Thus, whereas the earlier agreement was between the CHIC clinical partners and the CHIC infrastructure (represented by the CDP), the new agreement is designed so that it may also be entered into by external third party clinical institutions that are interested in testing and helping to validate the CHIC models. As regards the other party, during the lifetime of the project this would, as before, be the CDP acting on behalf of the CHIC consortium. However, the agreement is flexible in this regard, by referring to the party as the CDP/CHIC administrator. This caters for the point that for exploitation purposes after the end of the project, another legal entity, either an existing partner of the project or a third party that acquires the CHIC infrastructure could take over this role. This model agreement is appended to the present Deliverable in Appendix 3.

4.4.2 CHIC End User Agreements

As part of the CHIC first iteration data protection framework, end user contracts were concluded between the partners – in particular the modelers - using clinical data in the CHIC project and the CDP. These contracts provide for the protection and security of the data that the users need to access and process to carry out their work in the project. Thus, while the CDP is responsible for ensuring state of the art technical security of the data within the CHIC infrastructure (an obligation, which as noted above, it has subcontracted to a data security expert, Custodix), the data users for their part must ensure that appropriate technical and organisational measures are implemented within their own processing environment. They are also prohibited from disclosing any clinical data to parties outside the project, or from engaging in steps, such as matching of data sets, that could potentially identify individual patients from the data. To ensure strict compliance, the liabilities under data protection laws

¹⁵ See the Article 29 Working Party, Opinion 1/2010, p. 15.

¹⁶ As to these, see part 4.3 above.

for negligent violation of these obligations are adverted to, and an indemnity from a party in breach required to the benefit of other project partners. In addition, a penalty clause is proposed in the agreement that foresees a set amount of liquidated damages that the party may be required to pay to the consortium in such circumstances.

With reference to the stage of clinically validating the models, which is the subject of the second iteration data protection framework, these agreements will continue to govern the duties and liabilities of the modelling partners until the expiry of the project, when the agreements also terminate. This reflects the point that, for the modelers, there is no material change in position, in particular in the nature of the data they access and use, over the previous model-building phase, involving the processing of securely de-identified retrospective clinical datasets. Thus, as described in part 4.1, they will access and use data from the research domain to help them adapt and refine the performance of the models. The data in question comprises the outcome predicted for a given (securely de-identified) patient and the actual outcome observed by the clinician, plus (securely de-identified metadata as to the way the clinician executed the model). The data protection and security obligations of the modelers when handling this data are of a piece with their obligations in respect of the retrospective clinical data, as addressed in the existing data user agreements.

Post the expiry of the project, a new set of contracts will be required to cover ongoing processing by the modellers of this nature. Here the other party to the contract will be the legal entity that succeeds the CDP as data protection coordinator and administrator of the infrastructure. Whereas during the lifetime of the project the modelers have been limited to partners within the CHIC project, these agreements will also allow for the potential involvement of third party modelers, who wish their models to be tested by clinicians (and receive performance feedback data) via the CHIC infrastructure. A model CHIC end user agreement drafted for this purpose is appended as Appendix 4.

4.4.3 CHIC Data Security Agreement

In accordance with the first iteration data protection framework, the CDP is responsible for the security of data processing within the CHIC project. However, the technical security aspects of the framework, have been delegated by the CDP to the security expert partner in CHIC (Custodix), and there is a long-standing contractual relationship under which Custodix has successfully fulfilled the same function in respect of other projects (including ACGT and P-medicine). In the first iteration framework, the relevant respective obligations of the CDP and Custodix have been governed by the TTP agreement, which sets out the conditions and obligations under which Custodix will deploy state of the art security measures to protect the CHIC infrastructure, and also act as a trusted third party (TTP). The latter role reflects the fact that all clinical data transferred to the CHIC infrastructure in model development phase was subject to secure de-identification, including a second pseudonymisation¹⁷ by Custodix using dedicated state of the art pseudonymisation software (CAT), in which the initial data provider's pseudonym is replaced by a second pseudonym. At the same time, Custodix as the TTP has retained a key (linking the first and second pseudonymisation codes) to allow in exceptional circumstances for re-identifying a given patient, in particular if the project generates new information of specific importance for that patient's treatment.

In the context of the validation of the models where the clinicians executing the models will process data in single-pseudonymised form and know who the relevant patients (in their care) for whom the model provides a given prediction, this trusted third party function will not be required. However, at the subsequent stage, when modelers utilise the data in the same securely de-identified form as earlier when building the models, the possibility of their uncovering potentially significant information for an individual patient remains. Accordingly, the TTP mechanism in the first iteration agreement is duplicated in the second iteration

¹⁷ The first pseudonymisation was carried out by the clinical data provider itself.

agreement. Similarly, in other respects it remains essential that the CHIC infrastructure continues to incorporate stringent state of the art data security safeguards. As laid down in Article 17 of Directive 95/46/EC (and reiterated in Article 32 of the GDPR 2016/679), such safeguards need to be proportionate to the sensitivity of the data and risk of harm if a security breach occurs.

In common with the other first iteration agreements, the existing TTP agreement is due to terminate when the project expires in March 2017. As regards any clinical model validation that might begin prior to this date, it is submitted that the present obligations on Custodix are already phrased in a sufficiently flexible way (notably by reference to the above proportionality requirement in Article 17) to allow for any necessary adjustments to the project security architecture, e.g. to implement separation between the (personal) data in the clinical data repository, and the (securely de-identified) data in the research domain. Post the expiry of the project, it is evident (unless an entity takes over the CHIC infrastructure, which has sufficient in-house data security expertise) that a similar new agreement will be needed. Such an agreement (now called the CHIC Data Security Agreement), and between the (future) 'CHIC Platform Administrator' and the 'CHIC Data Security Provider' is appended as Appendix 5.

4.4.4 Further regulatory requirements for validation of models

4.4.4.1 Ethics Committee Approval for Clinician Validation

As noted in part 4.3 above, the assumption is that, during the CHIC model validation process, patients will not be exposed to invasive or burdensome processes that they would not otherwise be. This reflects the intention that the models will be deployed as an aspect of standard clinical practice – requiring them to be usable as a clinical information support tool by clinicians in possession of the standard kinds of data captured during patient diagnosis. However, if in some cases validation will in fact require non-standard procedures, so that patients are exposed to additional procedures, then the validation process may be regarded as a form of low interventional medical research: in that case, each clinical institution engaging in the validation would be required first to apply to its responsible ethics committee for prior approval (based on the submission of a CTP, including the model patient consent form).¹⁸

Nonetheless, even where the patient is not subject to any additional clinical procedure during the model validation, there remains in our view a strong argument that the clinical institution should seek the advice and approval of its ethics committee. This is because there is a risk that, during the validation process (as described under 4.1 above) a potential risk to the patient's interests may arise. This is that, through knowing the prediction of the model as to what the treatment outcome for the patient may be, the clinician may be led into a kind of observer bias in interpreting the patient's data and/or be unconsciously disposed to take treatment decisions on that basis. Admittedly, clinicians, as trained professionals, may be expected to guard themselves against such risks. Nonetheless, without detailed knowledge of the context in which a clinician will be deploying the model in a specific case, such a risk cannot be excluded a priori. This is the more so, given that what is at issue is a form of observational research, where patient data is processed not directly for use in the patient's care (the standard therapy is intended to be used, regardless of what the model predicts). In this regard, the model CHIC clinical user agreement in Appendix 3 contains an obligation on the clinical institution to obtain ethics approval before validating the models.

¹⁸ WMA Declaration of Helsinki (2013 revision), articles 22 ff.

4.4.4.2 Potential Medical Device Regulatory Approval

A second question concerns the application of the medical devices regime at the point of commencing clinical validation of the models. As examined in Deliverable D4.4 (M36),¹⁹ *in silico* models will very likely qualify as a 'stand-alone software' medical device within the definition of the EU Medical Devices Directive 93/42/EC ('MDD'), which regulates the testing, certification and post-marketing surveillance requirements in Europe for medical devices for human use. As such they would be subject to a set of pre-marketing testing and certification requirements, leading to the award of a CE mark as a precondition for being lawfully placed on the European market. According to the MDD, Annex X, the evaluation that the device satisfies the essential requirements should be based upon 'clinical data', showing the device does not pose undue safety risks to users, and also that it performs in the way the manufacturer claims. For the purpose of gaining such, the device may already be used in clinical practice as part of a 'clinical investigation' without yet bearing a CE mark.²⁰

At the same time, it remains uncertain in the context of the CHIC model validation at what point exactly 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 such an investigation; as noted in D4.4, the implications of it being so, are that the clinicians concerned would come under formal clinical data gathering duties, and their clinics would need to institute mechanisms to protect 'observer neutrality', such as by entrusting observations of the real patient to separate clinicians to those aware of the hypermodel's prediction. The details are themselves subject to current reforms to the medical devices regime, with changes likely also to the rules on clinical investigations.

In fact, for the present it is arguable that the sporadic 'test-driving' of the models by the CHIC clinicians, involving observations made with small numbers of patients should not be treated as a clinical investigation under the MDD. (By implication, the experience of the clinicians in using the models, and their observations concerning the accuracy or otherwise of the models would not at this point carry probative force as 'clinical data' for that purpose.) Later, once the predictions made by the models are sufficiently precise, proper clinical testing under the MDD would then be called for; however, this stage will not be reached in the project lifetime.

Nonetheless, given the uncertainty noted above regarding the application of the MDD on this point (requiring a fact-specific assessment by the relevant MDD regulatory authority) it is our view that clinical institutions wishing to test the CHIC models should consult with their respective national authorities to check that the latter share the view that what is proposed does not yet engage the rules under Article 15 MDD. A requirement to this effect is thus included in the model CHIC clinical user agreement. It is also important to note that, irrespective of whether the MDD is found to apply, the need for the clinical institution to seek the approval of its ethics committee to the validation process (as noted in 4.4.4.1) will remain.

4.4.4.3 Future involvement of Data Protection Authorities under the terms of the new General Data Protection Regulation

As noted earlier, and discussed in more detail in Deliverable D4.4,²¹ the current Data Protection Directive 95/46/EC will be replaced in 2018 by the new General Data Protection Regulation 2016/679. This will bring a number of changes to the regulatory landscape that governs data processing operations involving sensitive data, including the health data in

¹⁹ See deliverable D4.4, Whitepaper, Recommendations for an amended European legal framework on patients' and researchers' rights and duties in E-health related research, Chapter 5.

²⁰ MDD 93/42/EC, Article 15.

²¹ Note 19 supra, Chapter 4, section 3.

CHIC. A key new element in this regard is the checking/approval of particular processing operations by national supervisory authorities. Thus, under Article 37 projects that utilise large amounts of sensitive data will be required to install a data protection officer to monitor data protection compliance. Moreover, it is likely the processing of health data for research will qualify as a processing operation of ‘specific risk’ under Article 35, triggering the need for a Data Protection Impact Assessment to consider the risks to data subjects and appropriate mitigation strategies; in most cases prior checking will then be required by a supervisory authority under Article 36. While these measures do not formally take effect until 28 May 2018, it will be important for parties concerned with the ongoing validation of the CHIC models to be aware of and prepare for these requirements in good time.

5. CHIC Data De-Identification Process

5.1 Introduction

This chapter focuses on the documentation of the data de-identification process during the lifespan of the project, the technical implementation of the de-identification tools used in CHIC and the role of the CDP as controlling unit. Due to the sensitive nature of the data involved however, it should be noted that this report will not go into granular details of all techniques used and will not describe the content of the real data sets.²²

5.2 De-Identification of data sets – a context related undertaking

De-Identification is a process by which a data custodian alters or removes identifying information from a data set, thereby making it harder for users of the data to determine the identities of the data subjects,²³ enables sharing of data for research purposes whilst mitigating the risks for the concerned individuals.²⁴ Although the scope of application of the Data Protection Directive 95/46/EC applies only to personal data, there are different approaches in the Member States as to when data can be regarded as anonymous in a legal sense.

De-identification of data sets is not only about erasing or generalizing of (quasi-)identifiers. Research has shown how difficult it is to create “truly” anonymous data whilst retaining as much information as required for research purposes. Therefore, it is of importance to consider the context in which data is processed and the possibility of control to prevent re-identification. This is to lower the risk of re-identification to a negligible level,²⁵ e.g. by introducing measures such as sharing the de-identified data with researchers through a secure portal with strict access control and by concluding a data sharing agreement between the transferor and the transferee where the transferee commit not to re-identify the data subject or perform analysis outside of the agreed research request or give access to the data to a third party.²⁶ This has the consequence that depending on the specific context and the level of control, different de-identification measures might be needed in order to assess whether the risk of re-identification is negligible. For example public data sharing most probably does require stricter measures directly applied to the data.²⁷

In CHIC data is only shared in a secure and shielded environment through a sophisticated data protection and data security framework. As discussed in Chapter 4 above, this is also being extended in a suitably adapted second iteration to the stage of clinically validating models. Key elements of the framework are:

- The two times pseudonymized data is shared with researchers through a secure Clinical Data Repository. Download and upload of data is strictly controlled by the CDP.
- Additionally, there are contracts in place which establish legal relationship between providers of data and the controlling unit CDP and contracts between the CDP and the end-users who want to use the data for research, namely:

²² The data gathered in CHIC is further described in the CHIC WP3 deliverable D3.1: Report on scenarios and data from defined patients.

²³ Simson Garfinkel, “De-Identification of Personally Identifiable Information”, DRAFT NISTIR 8053.

²⁴ See Art. 29 WP, Opinion 05/2014 on Anonymization Techniques, adopted on 10.04.2014, p. 3.

²⁵ Art. 29 WP, Opinion 4/2007 on the Concept of Personal Data, adopted on 20.06.2007, p. 15.

²⁶ Phuse De-Identification Working Group, De-Identification Standard for CDISC SDTM 3.2, version 1.01, 20.05.2015, http://www.phuse.eu/Data_Transparency_access.aspx, (accessed 15.06.2015), Introduction page; Forgó et.al., Ethical and Legal Requirements for Transnational Genetic Research, 2010, pp. 91-112.

²⁷ Phuse De-Identification Group, op. cit.

- i) **Data Transfer Agreement:** This agreement is concluded between the CDP and the healthcare organization/hospital delivering patient data. The agreement for example entails an obligation for the transferor to pseudonymize the data and to assure the quality of pseudonymisation by implementing the Custodix Anonymisation Tool (CATS) or any other state of the art pseudonymisation tool recommended or accepted by the CDP; and
- ii) **Contract on Data Protection and Data Security within CHIC:** This agreement is concluded between the CDP and all end-users of CHIC doing research on the provided data. It entails several obligations for the transferees, such as not to re-identify data sets, not to disclose data to any other person unless in pursuit of their duties as detailed in the contract underpinned by a penalty clause.²⁸

5.3 Role of standards

While national legislation in Europe does not provide any specific requirements for effective anonymization, there are a number of guidelines available from public and private institutions which can assist data controllers in their de-identification tasks, such as:

- Art. 29 Working Party, Opinion 05/2014 on Anonymization Techniques;²⁹
- ICO, Anonymization: Managing Data Protection Risk Code of Practice;³⁰
- ISO/TS 2537, Health Informatics – Pseudonymization;
- Phuse De-Identification Working Group, De-Identification Standard for CDISC SDTM 3.2;³¹
- TransCelerate Biopharma Inc., Data de-identification and Anonymization of Individual Patient Data in Clinical Studies – A Model Approach;³²
- NIST, De-Identification of Personally Identifiable Information³³
- A De-Identification Strategy Used for Sharing One Data Provider's Oncology Trials Data through the Project Data Sphere® Repository;³⁴
- Novartis Global Data Anonymization Standards;³⁵
- HITRUST De-Identification Framework.³⁶

Though these standards can provide valuable assistance for de-identification of data sets, it however remains the responsibility of the data controller to ensure the anonymous state of the data. This means that a constant review of the techniques is required in order to keep the anonymization up to date using a state of the art.

²⁸ For further discussion and a draft of these contracts, see part 4.4 above and Appendices 3-5 below.

²⁹ Art. 29 Working Party, Opinion 05/2014 on Anonymization Techniques, adopted on 10.04.2014.

³⁰ ICO, Anonymization: Managing Data Protection Risk Code of Practice, November 2012, <https://ico.org.uk/media/1061/anonymisation-code.pdf> (accessed 15.06.2015).

³¹ Phuse De-Identification Working Group, De-Identification Standard for CDISC SDTM 3.2, op. cit.

³² TransCelerate Biopharma Inc., Data de-identification and Anonymization of Individual Patient Data in Clinical Studies – A Model Approach, <https://www.transceleratebiopharmainc.com/wp-content/uploads/2015/04/Data-Anonymization-Paper-FINAL-5.18.15.pdf> (accessed 15.06.2015).

³³ Draft NISTIR 8053, op. cit.

³⁴ Project Data Sphere, LLC, A De-Identification Strategy Used for Sharing One Data Provider's Oncology Trials Data through the Project Data Sphere® Repository <https://www.projectdatasphere.org/projectdatasphere/html/resources/PDF/DEIDENTIFICATION> (accessed 15.06.2015).

³⁵ Novartis Deutschland GmbH, Novartis Global Data Anonymization Standards (accessed 15.06.2015).

³⁶ HITRUST, HITRUST De-Identification Framework, <https://hitrustalliance.net/de-identification/> (accessed 15.06.2015).

5.4. De-Identification approach in CHIC

5.4.1 CHIC data de-identification flow

Within CHIC data is de-identified through a first round done at the client by the data source and a second by the Custodix' Pseudonymization Services.

By contract, it is the responsibility of a data provider to de-identify a data set (first round) before exporting it to the CHIC research domain. The data uploader can use any existing data de-identification tool. CHIC provides a default implementation with the Data Upload Tool and the CATS (Custodix Anonymisation Tool Services).

The CATS server, hosted by Custodix, is responsible for the second de-identification round. This service will encrypt the patient pseudonyms with a key held by a Trusted Third Party so that reversal and re-identification is not possible without the Trusted Third Party's intervention.

When the data provider makes use of the default CHIC de-identification tools, the CDP (Centre of Data Protection) operatives will use their privacy expertise to assist the data source in defining in their datasets the identifiers, quasi-identifiers and sensitive fields. The CDP will also suggest the transformations to perform and how to configure these for those mapped fields in a privacy profile executable by CATS.

Through the CHIC Upload Tool (developed by FORTH) and CATS, these data files are then de-identified by executing the previously defined privacy profiles and uploaded to the CHIC de-identification service for a second de-identification round.

Before a dataset can be actually imported in to the clinical data repository, it is the CDP's responsibility to validate the dataset and ascertain that the file has indeed been sufficiently de-identified. If rejected, the CDP will either contact the data source or Custodix (responsible for the second de-identification round) so that corrections be made and resubmit the data for approval. Once approved, the data can be released to the CHIC clinical data repository.

5.4.2 Using Syntactic De-Identification Techniques

As a result of the needs of the project, syntactic de-identification techniques have been used. These techniques attempt to de-identify a dataset by performing transformations on a data set based on field types.

For each dataset we started by defining the identifiers, quasi-identifiers and sensitive fields (or attributes)³⁷.

- **Direct Identifiers** are fields that clearly and uniquely identify individuals (such as patient number, social security number, address, name)
- **Quasi-identifiers** are fields whose values when combined, linked with other records, public information or background knowledge can with a high probability identify an individual (e.g. zip-code, birthdate, gender, weight, length, race, specific findings, adverse events)
- **Sensitive fields** are those fields that in case of a breach and re-identification would harm the patient in terms of self-esteem, loss of income (disease diagnosis, salary ...), insurability, employability or reputation³⁸.

Once fields have been marked, operations (or transformations) to be performed on these fields are defined.

³⁷ Ninghui Li, Tiancheng Li, Suresh Venkatasubramanian. t-Closeness: Privacy Beyond k-Anonymity and l-Diversity. ICDE, IEEE 2007, pp. 106-115.

³⁸ See e.g. Phuse De-Identification Working Group, De-Identification Standard for CDISC SDTM 3.2, version 1.01, op. cit., Definitions page.

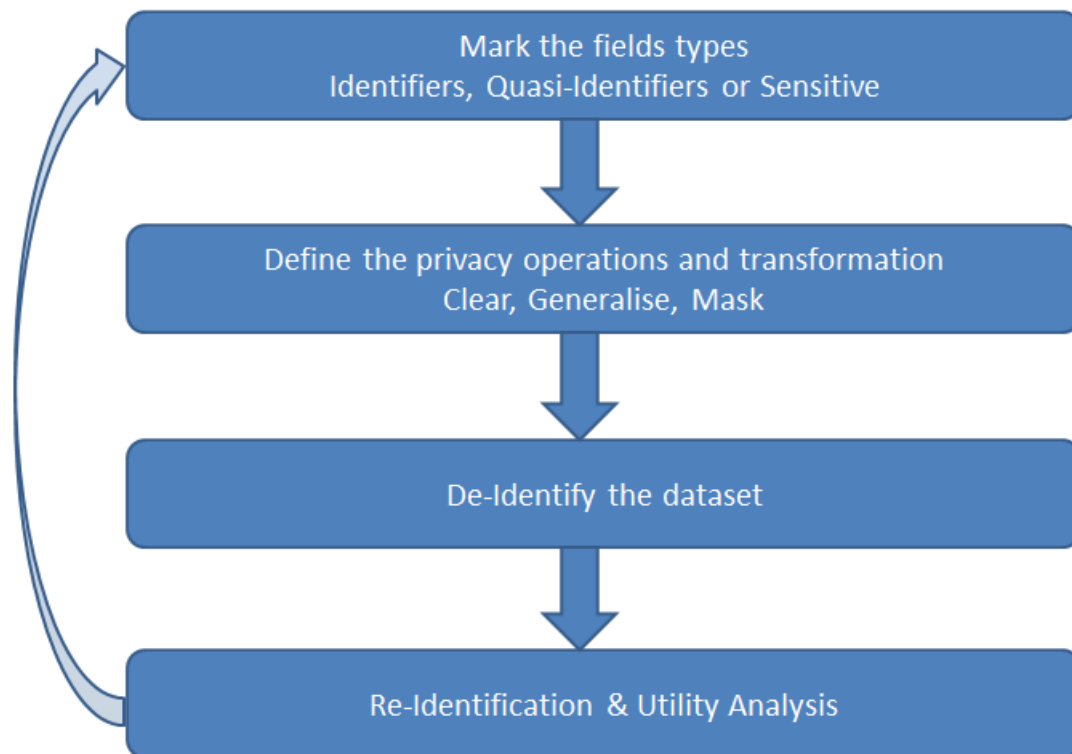


Figure 2 Syntactic de-identification

This results in a privacy profile which defines how a given file should be de-identified. After execution of the profile the de-identified data should be validated on:

1. the re-identification risk; whether the chance of re-identifying a record is sufficiently low and meets a predefined threshold.
2. its utility; whether the data is still useful for the researchers to do their analysis.

If the data does not satisfy this validation, the procedure is repeated, until the de-identification is satisfied. Once validated and approved the data can be released.

5.4.2.1 Direct identifiers

Direct identifiers directly identify an individual and usually do not contain any useful information. These are suppressed and possibly replaced by either a random (anonymous) identifier or pseudonym. When using random identifiers subjects from different datasets cannot be associated anymore to each other. Pseudonyms either calculated from the identifiers or randomly generated and stored in a linking table, allow subjects from different datasets to be linked to each other and to be re-identified by the owner of the calculation key or linking table. Using a pseudonym – although re-identification is in principle possible – does not prevent to have an anonymized data set if certain circumstances are given such as for CHIC. The CHIC security framework where data is also pseudonymized, keeps the risk of re-identification to such a negligible level that the data within that framework can be regarded as anonymous in a legal sense.

5.4.2.2. Quasi-identifiers

Quasi-identifiers, although not directly identifying by themselves, can be used to identify an individual when combined with each other, linked to other public records or background knowledge. Therefore to avoid the disclosure of sensitive information, these quasi-identifiers are also de-identified within our framework. Quasi-identifiers such as weight or age generally contain information important for analysis. When transforming the quasi-identifiers a right

balance should be found in limiting disclosure risk while maximising the usefulness of the data. For this, different transformation techniques exist:

- Through **generalisation**^{39, 40} the values of a given field are substituted with more general values. Generalisation models data in generalisation hierarchies where the leaves are the actual values of the field. By moving to a higher level the precision of a value is reduced. A de-identification algorithm needs to deal with this hierarchical nature of such fields.
- **Suppression**⁴¹ removes data. It can be applied at the level of a row by removing a whole record, at the level of a field by clearing all values of an attribute or at the level of a cell by removing the value for a specific field and record.
- **Global recoding** aggregates the values of a field causing several values of that field to be collapsed into a single one.
- **Post-randomisation Method** (PRAM) protects categorical attributes (e.g., blood type) from disclosure. It uses a known probability mechanisms, the values of a categorical attribute are changed to a new values, which may or may not be different from the original. It basically deliberately misclassifies a field, wherefore it will be difficult to identify records (with certainty) as corresponding to certain individuals. Since the probability mechanism is known, characteristics of the true data can still be estimated from the de-identified data.
- **Microaggregation**⁴² aggregates records into groups. Instead of releasing the actual values of sensitive attributes, the mean of the group to which the observation belongs is released. The confidentiality of individual data subjects is protected by ensuring that each group has at least a minimum number of observations.
- **Top- and bottom coding** are particular global recoding techniques. Top coding on numerical fields groups together all top values. It sets an upper limit on all values of that field. Bottom coding groups together all bottom values.
- **Slicing**⁴³ partitions data vertically into groups of correlated fields. It then horizontally partitions the data into groups of records. Within each group of records, the rows in each field are randomly permuted to hide the linking between the different fields. Generalisation is in addition often applied.

Other, lesser discussed, techniques exist such as adding-noise, data-swapping (specific for categorical data) and re-sampling (specific for numerical data).

5.4.3 Overview of CHIC de-identified datasets

During the existing period to date of CHIC, several datasets from multiple partners (in diverse formats) have been successfully de-identified and approved for sharing (see Table A for an overview).

³⁹ Kristen LeFevre, David J. DeWitt, Raghu Ramakrishnan. Incognito: Efficient FullDomain KAnonymity, 2005.

⁴⁰ Xiaoxun Sun, Min Li, Hua Wang, Ashley Plank. An efficient hash-based algorithm for minimal k-anonymity, 2008.

⁴¹ V. Ciriani, S. De Capitani di Vimercati, S. Foresti, and P. Samarati. k-Anonymity

⁴² Stephen Lee Hansen, Sumitra Mukherjee. A Polynomial Algorithm for Optimal Microaggregation, 2003

⁴³ Tiancheng Li, Ninghui Li, Jian Zhang, Ian Molloy, Slicing: A new approach to privacy preserving data publishing. 2009

Institution	Kind of data	Data format	# patients	# data	Status
USAAR	Nephroblastoma Initial Dataset	CDISC ODM XML, DICOM, CSV, miRNA (miniml)	86	Clinical data (+/- 40 fields) DICOM data on 21 patients (+/- 1600 files, 500 mb) miRNA data on 66 patients	Shared Shared Shared
USAAR	Nephroblastoma Validation Dataset	CDISC ODM XML, DICOM, CSV, miRNA (miniml)	61	Imaging, miRNA and clinical data. (+/- 70GB of DICOM data)	October 2016
USAAR	Lung	CSV, DICOM, miRNA (miniml)	100	Clinical data DICOM data on +/- 48 patients for a total of +/- 47000 DICOM series/sets (64GB) miRNA data on +/- 20 patients	Shared Shared Shared
KUL	Glioblastoma Initial Dataset	CDISC ODM XML, DICOM, CSV	82	+/- 200 data fields per patient, 10 CRFs (eg. blood counts, questionnaires, vaccination, radiology, chemotherapy, medication) 3 to 21 time points per patient for a total of +/- 780.000 files with +/- 7000 DICOM series/sets (31 GB) Serum data (6 parameters over 4 time points) PBMC (12-13 parameters over 4 time points) Pathologies	Shared Shared Shared End of 2016 End of 2016
KUL	Glioblastoma Validation Dataset	DICOM, CSV	52	+/- 36 data fields per patient 3 to 21 time points per patient for a total of +/- 500.000 files with +/- 4500 DICOM series/sets (20 GB) Serum data (6 parameters over 4 time points) PBMC (12-13 parameters over 4 time points) Pathologies	Shared DICOM expected 4 th quarter of 2016 Shared End of 2016 End of 2016
UNITO	Prostate	CSV, miRNA	1161	1161 patients with complete follow-up over 5 years. +/- 25 relevant fields per patient miRNA available for 10 patients	Expected 4 th quarter of 2016

Table A: Overview of de-identified, approved and shared datasets⁴⁴

5.4.4 CHIC Data de-identification process

Through a sample clinical tabular data file (such as CSV) this section will explain the CHIC data de-identification process.

The data upload and de-identification process is initiated by the physician who wants to upload a clinical data file to the CHIC clinical data repository for research purposes.

Hospital Identifier	Name	Birthday	Gender	Length	Weight	Date of Diagnosis	Diagnosis

Figure 1: Example Clinical Data Table

In collaboration with a privacy expert from the Centre of Data Protection the clinician starts by identifying the direct identifiers, quasi identifiers and sensitive fields in the dataset.

k

Identifiers		Quasi Identifiers					Sensitive Data
Hospital Identifier	Name	Birthday	Gender	Length	Weight	Date of Diagnosis	Diagnosis

Figure 2: Example Clinical Data: field types marked

Once all fields have been typed, the clinician (again in collaboration with a privacy expert from the Centre of Data Protection) defines the transformation to be performed:

- Replace the hospital identifier by a reversible pseudonym and clear all other direct identifiers.
- Randomise the day of birth by shifting it with a different random interval for each patient. All other dates belonging to the same patient should be shifted accordingly to retain the time interval.
- Top and bottom code length and weight to remove outliers.

⁴⁴ As this is a public deliverable, for privacy reasons we cannot go into details on the processed data.

Identifiers		Quasi Identifiers					Sensitive Data
Pseudonymize	Suppress	Shift		Top & Bottom Code	Top & Bottom Code	Shift	
Hospital Identifier	Name	Birth day	Gender	Length	Weight	Date of Diagnosis	Diagnosis

Figure 3 Example Clinical Data: field types marked

After the de-identification strategy has been defined, a CATS privacy profile is created on CATS so that the CHIC Upload Tool can automatically de-identify the data during the upload process.

In CATS, field typing and transformation definition is done by mapping fields to variables, by optionally assigning a privacy type and by finally defining operations on the variables.

```
<dataMapping>
  <simpleVariable name="ID" columnName="Hospital Identifier" privacyType="privacy:identifier"/>
  <simpleVariable name="name" columnName="Name" privacyType="privacy:identifier"/>
  <simpleVariable name="gender" columnName="Gender"/>
  <simpleVariable name="length" columnName="Weight"/>
  <simpleVariable name="weight" columnName="Length"/>
  <simpleVariable name="dateOfBirth" columnName="Birth day" privacyType="privacy:date" sourceType="xsd:string" type="java:date" conversionParams="dateFormat (dd/MM/yyyy)"/>
  <simpleVariable name="dateOfDiagnosis" columnName="Date of Diagnosis" privacyType="privacy:date" sourceType="xsd:string" type="java:date" conversionParams="dateFormat (dd/MM/yyyy)"/>
</dataMapping>
```

Figure 4 Example of CATS field mapping of Clinical Data

```
<process>
  <forEachDataSource>
    <forEachDataSourceRow>
      <function ref="core:randomId">
        <forEachVariable itemName="identifier" privacyType="privacy:identifier" condition="identifier.updated=='false'">
          <function ref="core:clear">
            </function>
          </forEachVariable>
        <function ref="core:randomDate">
          <forEachVariable itemName="date" privacyType="privacy:date" condition="date.updated=='false'">
            <function ref="core:makeDateRelative">
              </function>
            </forEachVariable>
          </forEachDataSourceRow>
        </forEachDataSource>
      </process>
```

Figure 5 Example of CATS transformation definition of Clinical Data

```
<function ref="core:randomId">
  <argument name="entityIdentifier">
    <variable ref="ID"/>
  </argument>
  <argument name="outputLength">
    <constant type="java:int" value="20" valueType="xsd:string"/>
  </argument>
  <argument name="database">
    <variable ref="projectDatabase"/>
  </argument>
  <argument name="tableName">
    <constant value="patientIDLinkTable"/>
  </argument>
  <argument name="reusePreviouslyGeneratedAnons">
    <constant value="true" type="java:boolean"/>
  </argument>
  <returnValue name="ID"/>
</function>
```

Figure 6 Randomise Patient ID and store it in local linking table

```
<forEachVariable itemName="identifier" privacyType="privacy:identifier" condition=
"identifier.updated=='false'">
  <function ref="core:clear">
    <returnValue name="identifier" />
  </function>
</forEachVariable>
```

Figure 7 Clear all other unmodified identifiers

```
<function ref="core:randomDate">
  <argument name="entityIdentifier">
    <variable ref="ID"/>
  </argument>
  <argument name="database">
    <variable ref="projectDatabase"/>
  </argument>
  <argument name="tableName">
    <constant value="referenceDate"/>
  </argument>
  <argument name="reusePreviouslyGeneratedDates">
    <constant type="java:boolean" value="true"/>
  </argument>
  <argument name="minYear">
    <constant value="1900" type="java:int"/>
  </argument>
  <argument name="maxYear">
    <constant value="1905" type="java:int"/>
  </argument>
  <returnValue name="dateOfBirth"/>
</function>
```

Figure 8 Replace date of birth by a new random date between 1900 and 1905 (shift)

```
<forEachVariable itemName="date" privacyType="privacy:date" condition=
"date.updated=='false'">
  <function ref="core:makeDateRelative">
    <argument name="inputDate">
      <variable ref="date"/>
    </argument>
    <argument name="referenceDate">
      <variable ref="dateOfBirth" original="true"/>
    </argument>
    <argument name="newReferenceDate">
      <variable ref="dateOfBirth"/>
    </argument>
    <returnValue name="date"/>
  </function>
</forEachVariable>
```

Figure 9 Shift all other dates of the same patient with same shift interval

Once created this privacy profile is uploaded to the CATS privacy profile store and mapped to the data file through media type and CSV header names.

Profile details

Name:

Source type:

Profile selector

Profile selection phase:

Mime type:

CSV Headers:

Profile configurations

Profile type:

Processing phase:

Privacy configuration files

No file chosen

Profile keystores

Figure 10 Privacy Profile Configuration on CATS

In addition to the source's profile, Custodix created a profile for the second pseudonymization round on CATS. This profile is similar to the source profile but will encrypt the pseudonyms created at the source so that it cannot be reversed anymore without access to the encryption key held by the TTP.

Once all privacy profiles have been created, the clinician can go ahead with the data upload through the CHIC Upload Tool which will download the previously defined privacy profiles, de-identify the data file and upload it to the CHIC pseudonymization service (CATS) for the second pseudonymization round. Once fully processed, the file is held on CATS for verification and validation by a CDP operative.

The CDP's evaluation of data file is based on three approaches.

- Verification of the syntactic de-identification performed (this is the privacy profiles created by the clinician (first round) and Custodix (second round).
- Manual viewing of the de-identified dataset.
- Perform some automatic validation algorithms such as
 - Group size calculation based on the quasi-identifiers (k-anonymity⁴⁵)
 - Risk calculation

At this stage, the CDP describes in its evaluation whether relevant identifiers have been de-identified, and assess the risk of unauthorised re-identification. A decision could then be reached on the release of the data and subsequent upload to the clinical data repository. Where there are issues with the data, the CDP could refuse to release it until appropriate actions are taken to remedy the issues and/or reprocess the data.

⁴⁵ Khaled El Emam and Fida Kamal Dankar. Protecting Privacy Using k-Anonymity, 2008.

5.5 Custodix Anonymisation Tool Services (CATS)

5.5.1 Custodix PseudoEngine

The Custodix Pseudonymization Engine (Pseudo Engine) is implemented as a layered model. At the core, the Pseudo Engine Core implements the pseudonymization language, the privacy profiles and support for input and output data values. Around the core various extensions provide extra functionality to the pseudonymization engine such as privacy and transformation functions, data input and output stream (this allows the processing of data files) and data mapping capabilities for e.g. XML, CSV and DICOM. The wrappers finally provide a graphical user interface, management and execution capabilities to the engine.

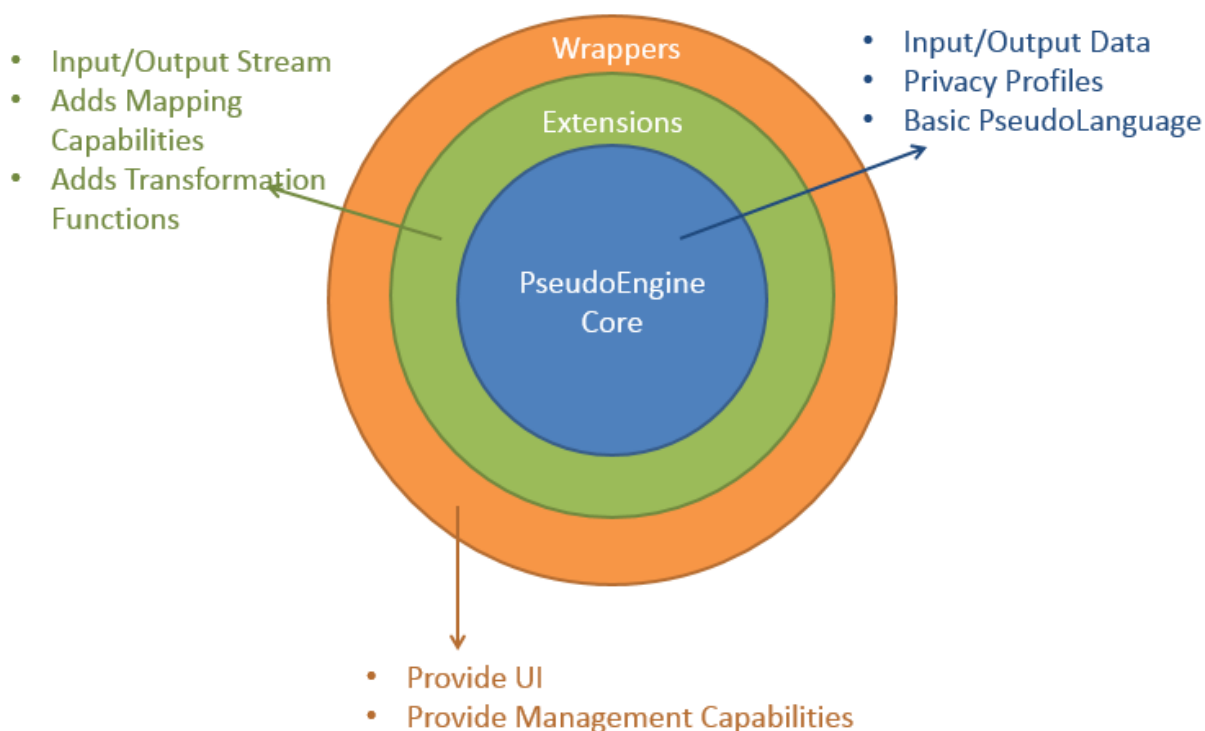


Figure 11 Pseudo Engine Model

5.5.1.1 Core PseudoEngine

The core pseudonymization engine takes as input a set of named variables and a privacy profile which defines how those variables should be processed. The output of the executed profile is a set of values.

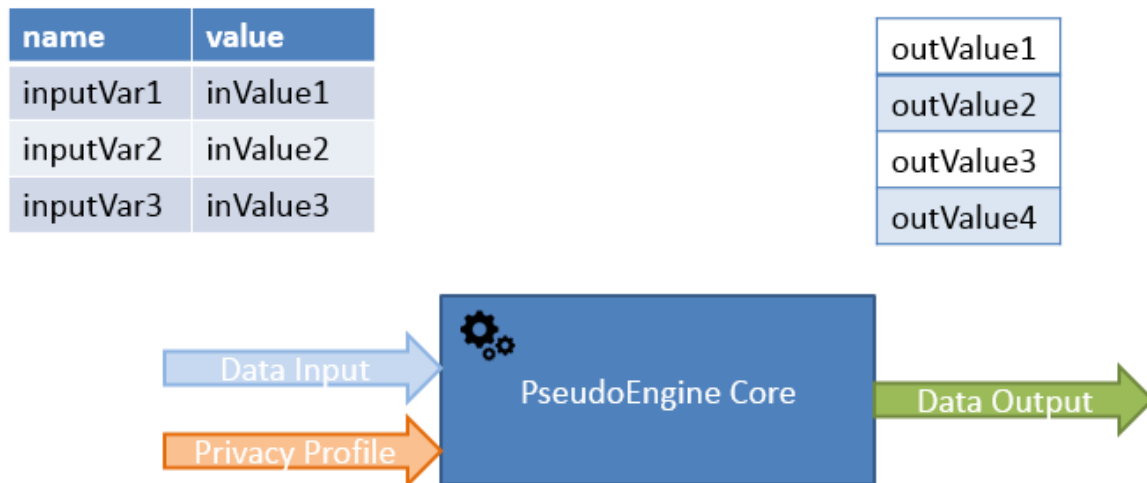


Figure 12 Pseudo Engine Core

5.5.1.2 PseudoEngine Data Mapping Extension

The data mapping extension extends the core functionality with support for streams both as input and output. A privacy profile now doesn't just define transformations on variables, but it also defines a mapping from the stream to named variables.

When executing a privacy profile on a data stream, the stream will be read in, parsed into variables and transformations applied on them. Once all transformations have been applied the variables are again outputted as a data stream. Supported stream types are for example CVS, XML, DICOM.

5.5.1.3 PseudoEngine Data Mapping Extension

The data mapping extension extends the core functionality with support for streams both as input and output. A privacy profile now doesn't just define transformations on variables, but it also defines a mapping from the stream to named variables.

When executing a privacy profile on a data stream, the stream will be read in, parsed into variables and transformations applied on them. Once all transformations have been applied the variables are again outputted as a data stream. Supported stream types are for example CVS, XML, DICOM.

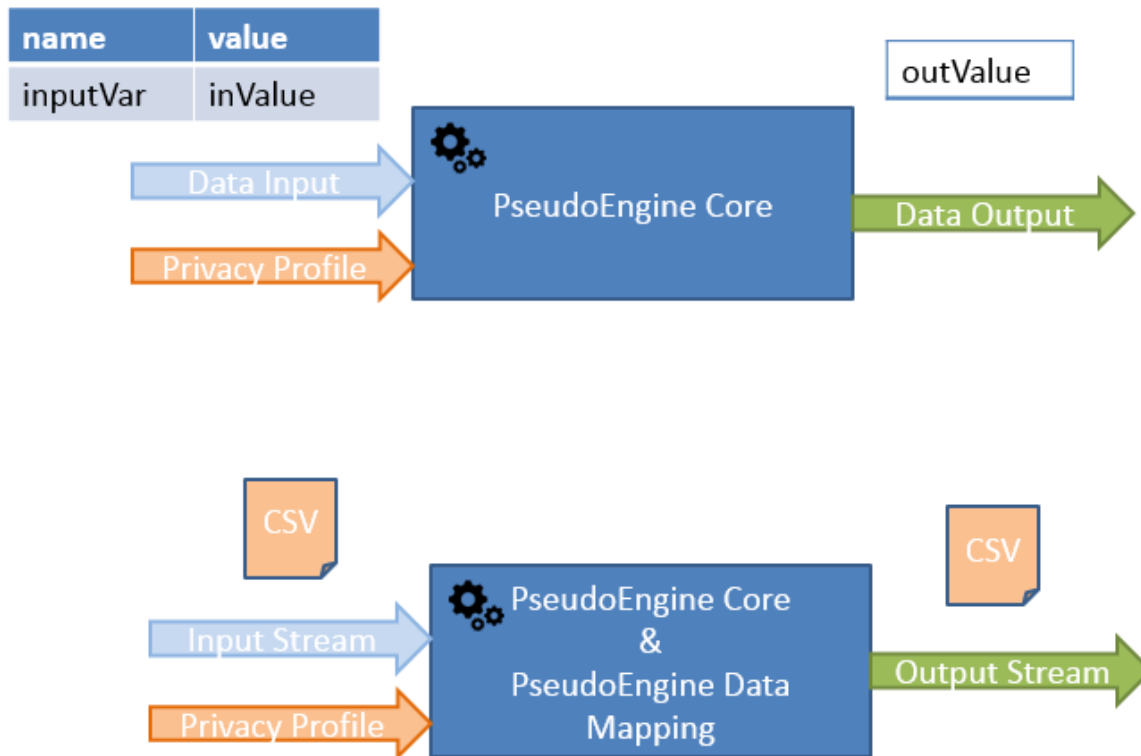


Figure 13 Data stream processing

5.5.1.4 PseudoEngine Function and Crypto Extension

The pseudo engine function extension provides implementation for various data transformation functions such as suppression, date generation, date randomisation (shifting over a random internal fixed for each patient), secure hashing and encryption.

5.5.1.5 Custodix Anonymisation Tool (CAT)

CAT is a pseudonymization engine wrapper that can be used for creating and testing simple privacy profiles. Cat is implemented as an eclipse-based client application. Privacy profiles created with CAT can then be used by other wrappers, such as command & upload tool, to process actual data files.

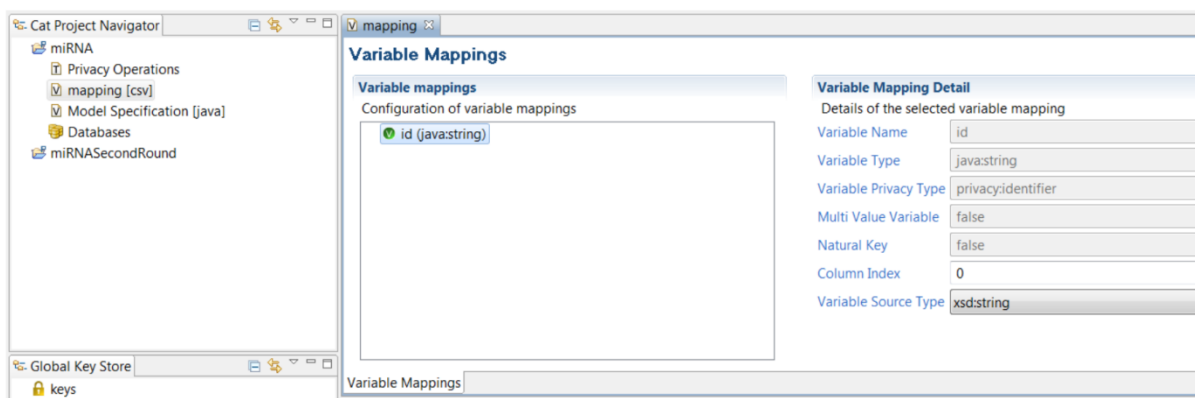


Figure 14 CAT Workbench Privacy Profile Creation

5.5.1.6 Custodix Anonymisation Tool Services (CATS)

The CATS wrapper is a service oriented evolution of CAT. CATS can be used to create/manage privacy profiles and to execute those profiles on data files. To accept files for processing and delivery of the de-identified result files, CATS provides various input and output interfaces such as web, SOAP and REST web services, sftp, email, file system, DICOM server.

Id	Username	File name	Date	State	Phase	Actions
2800280	f13b99ec-7799-4283-8e88-d399e27b893d	demo.xml	2015-06-09 10:10:57.0	TO_BE_REVIEWED	VALIDATED	Accept Reject Download Upload
2670268	bb86e4ca-048c-404c-90c0-7773ac22dc60	demo.xml	2015-06-05 11:51:49.0	DELIVERED	FINALISED	
2670267	bb86e4ca-048c-404c-90c0-7773ac22dc60	demo.xml	2015-06-05 11:50:50.0	DELIVERED	FINALISED	
2580258	bb86e4ca-048c-404c-90c0-7773ac22dc60	demo.xml	2015-06-05 11:17:18.0	DELIVERED	FINALISED	
2550255	bb86e4ca-048c-404c-90c0-7773ac22dc60	demo.xml	2015-06-05 09:15:53.0	DELIVERED	FINALISED	
2530253	bb86e4ca-048c-404c-90c0-7773ac22dc60	demo.xml	2015-06-04 16:10:29.0	DELIVERED	FINALISED	
2510251	bb86e4ca-048c-404c-90c0-7773ac22dc60	demo.xml	2015-06-04 16:04:40.0	DELIVERED	FINALISED	
2310231	f13b99ec-7799-4283-8e88-d399e27b893d	demo.xml	2015-04-21 13:10:51.0	DELIVERED	FINALISED	
2080209	bb86e4ca-048c-404c-90c0-7773ac22dc60	demo.xml	2015-04-09 16:57:54.0	DELIVERED	FINALISED	
2080208	bb86e4ca-048c-404c-90c0-7773ac22dc60	demo.xml	2015-04-09 16:56:51.0	DELIVERED	FINALISED	
2010201	bb86e4ca-048c-404c-90c0-7773ac22dc60	demo.xml	2015-04-09 15:52:02.0	REJECTED	VALIDATED	
1940194	f13b99ec-7799-4283-8e88-d399e27b893d	demo.csv	2015-04-09 08:51:14.0	ERROR	CLEANED	Reprocess Confirm

Figure 15 CATS uploaded/processed file overview

CATS processes files by selecting a correct privacy profile from its profile store and then executing that privacy profile on uploaded data through an embedded pseudonymization engine.

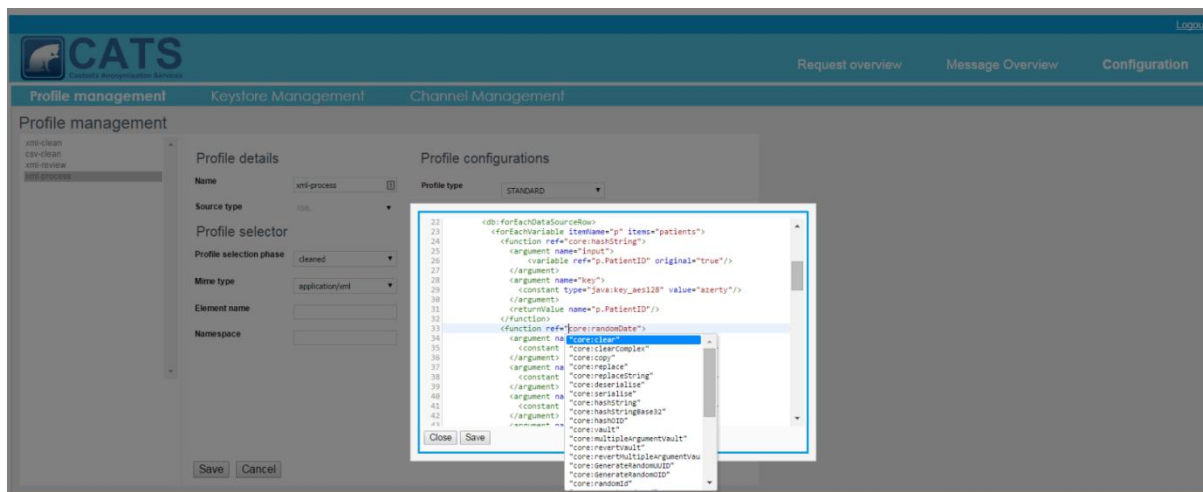


Figure 16 CATS Profile Editor

5.5.1.6.1 CHIC Upload Tool

The CHIC Upload Tool is an engine wrapper (with embedded pseudonymization engine for client side de-identification) through which a source can pseudonymize a data file and upload it through the CHIC pseudonymization services (CATS) into the CHIC clinical data repository.

File History Help															
Upload Upload results															
hsa-let-7c	hsa-let-7c*	hsa-let-7d	hsa-let-7d*	hsa-let-7e	hsa-let-7e*	hsa-let-7f	hsa-let-7f-1*	hsa-let-7f-2*	hsa-let-7g	hsa-let-7g*	hsa-let-7i	hsa-let-7i*	hsa-miR-1	hsa-miR-100	hsa-miR-100*
4.7233838...	4.7233838...	11.424745...	4.723383843	8.72331097	4.723383843	4.7233838...	6.544654879	6.359494822	6.5446548...	4.723383843	7.6196362...	7.7940808...	4.7233838...	6.703082941	4.723383843
9.2123146...	7.0120977...	10.790628...	8.418359249	7.3116601...	4.734239033	9.0024427...	4.166071497	3.658014102	7.5436428...	7.012097743	8.81697378	7.7202668...	4.3794021...	8.540595492	2.483322428
8.8013786...	7.0120977...	11.340131...	8.514525335	7.6638784...	5.704085169	9.0928976...	6.082630241	3.019118671	8.0648267...	6.424074833	9.0928976...	7.4837815...	3.0191186...	8.472305179	4.867946854
8.8521333...	5.8330899...	9.8696283...	5.676882024	6.5303767...	4.858986777	7.44902022	4.166071497	3.005617705	7.3502546...	7.726067798	9.1785572...	8.2790788...	3.7588338...	6.672544741	6.407500434
6.8930196...	5.1958141...	8.9008331...	6.205484368	4.5053585...	5.414088181	5.6949632...	3.728155796	3.955066435	4.8805781...	7.253825225	6.65863114	8.3403028...	3.7666300...	6.338605719	6.506744605
10.021368...	6.3483777...	11.087588...	8.172541757	7.2740231...	2.888684219	8.8521333...	7.372382042	2.888684219	8.2220050...	7.928780705	9.6440969...	8.5544899...	4.6391284...	8.320275587	6.857068942
3.0035947...	8.5377799...	9.50529086	5.379592981	3.6580141...	5.704085169	1.7083965...	4.076923526	2.557816333	5.2121560...	6.494878661	4.0769235...	8.29548852	4.8433042...	7.04485565	5.379592981
6.7924262...	5.36335669	10.099861...	8.733391746	7.2538252...	4.378129346	7.9778777	3.009236826	3.009236826	6.5977643...	6.597764334	7.6579320...	8.3574601...	3.0092368...	8.625059257	4.872032243
5.8668485...	6.6983376...	9.6606710...	8.087606125	7.11764331	3.762803205	8.3960311...	3.762803205	3.762803205	3.7628032...	5.866848585	6.9210741...	7.11764331	7.3383204...	5.866848585	5.36036786
7.3116601...	3.3766680...	9.4466940...	9.002442774	7.4863171...	5.998360546	8.3754153...	5.998360546	3.376668041	3.3766680...	7.311660123	7.4863171...	7.4863171...	3.3766680...	9.080829579	5.410654653
8.0876061...	6.71211177	10.5000806	8.685372759	7.6146722...	4.076923526	8.8325689...	6.999353586	5.870598994	8.1725417...	5.870598994	8.7333917...	6.71211177	7.4709164...	6.999353586	4.076923526
9.0808295...	3.2162851...	10.5000806	7.787379607	7.8199610...	6.387462072	8.9809862...	4.808509107	4.808509107	8.6354799...	6.051431035	9.5242612...	6.0514310...	5.2167420...	6.637809265	4.808509107
8.3403028...	5.0466706...	10.021368...	8.254187269	7.1785316...	5.379592981	8.66239472	6.93335665	5.379592981	7.8237219...	6.557006216	8.7434725...	7.8550221...	6.0826302...	8.287283699	6.082630241
9.2657434...	2.2635218	11.087588...	8.66239472	7.7030490...	4.844067734	9.0808295...	3.765135589	3.379927801	7.9107799...	6.355715272	8.72331097	7.9699273...	5.0873465...	7.278415973	6.016831979
6.3644987...	4.0913690...	9.8266503...	4.091369032	4.0913690...	4.091369032	4.0913690...	6.096373112	4.091369032	4.0913690...	6.364498772	4.0913690...	8.7026903...	4.0913690...	10.00053205	4.091369032
2.45721951	6.4570602...	8.2060138...	7.277263767	4.8589867...	5.413542899	5.4135428...	5.106951239	2.45721951	2.45721951	7.740924565	5.4135428...	9.5885594...	5.4135428...	6.007125687	7.044486283
2.53699564	5.2543124...	2.8826390...	4.87674568	1.8346213...	3.851486544	4.7336144...	5.087346528	4.082973119	1.8346213...	7.420548013	1.8346213...	9.7576201...	3.5045446...	6.3934567	5.414633463
8.2271846...	2.4691242...	10.218166...	8.66239472	6.0889343...	5.641267736	7.7409245...	4.091369032	5.904890294	7.1285563...	6.088934313	7.0444862...	8.4129045...	2.4691242...	8.832568945	6.342809363
8.7130006...	2.6388251...	10.751009...	8.713000666	7.0444862...	6.348377773	9.0121917...	3.845919257	5.222737492	7.8847575...	7.271934836	8.81697378	8.3960311...	4.1660714...	5.999991865	6.348377773
9.0219408...	3.4581448...	11.284334...	8.320275587	6.9511408...	3.458144882	9.4182591...	6.451356491	3.458144882	7.9937501...	7.290836182	8.9666085...	8.4383027...	3.4581448...	7.290836182	6.451356491

5.5.1.6.2 CLI Tool

The CLI command line tool wrapper is a command line application that allows the execution of a profile on a data file through a command line interface.

```
./CLI "--dbURL=jdbc:mysql://xxxxxx/pmed2015" "--dbDriverClass=com.mysql.jdbc.Driver"
--dbUsername=xxxxxx --dbPassword=xxxxxx -Kkeys=pmedKeys.catbks -Pkeys=xxxxxx
-c process_dicom.catop -c process_dicom.catmap -Iin=dicom -Tin=DICOM -Oin=dicom-out -LGERMAN
```

Figure 17 CLI command line call

5.5.2 Custodix Privacy Language

5.5.2.1 Privacy Profile

A privacy profile is a script which defines how an input file should be de-identified. A privacy profile is created in an XML language called the Custodix Privacy Language.

```
<pseudonymisation xmlns="http://www.custodix.com/pseudonymisation/language">
```

Data Mapping

Custom Function Definitions

Operations or Data Protection Profile

```
</pseudonymisation>
```

A privacy profile is structured in three sections.

- The mapping of a data file (which fields are identifiers, quasi identifiers and sensitive).
- A set of custom defined privacy operations/functions.
- The actual privacy operations to be executed on the mapped data.

A privacy profile is executed by the Pseudo Engine. The engine will start by reading in and parsing the XML privacy profile. From that it will create an executable syntax tree which will be evaluated by the Pseudo Engine. Such a syntax tree is build up by evaluable nodes. During evaluation each node will be assigned a value. The value of a node is typically calculated by performing an operation with as input the values of its children. Some nodes can also update values from the context. The value the root node evaluates to is the result of the executed privacy profile.

Through the data mapping extension the engine can in addition process streams (both as input and as output). In this case the engine can have as result, a result value and a processed data file written to the output stream.

5.5.2.2 Context and variable typing

The core the pseudo engine takes as input a privacy profile and a set of named variables. Once execution finished the engine will have as result a set of result values. Engine variables are not just named but are also aware of their type (e.g. java:string for a characters string, java:int for an integer number).

Upon execution the engine will create a series of nested execution contexts. These contexts contain named and typed variables. At the root sits the Application Context (also referred to as the global context). This context typically contains some global variable and function definitions. As child of the Application Context the engine will create the Execution Context. This context contains all the variables the engine is working with. The execution context is initialised with the named variables passed to the engine at start-up. Figure 18 below shows a very simple example where two variables are passed to the engine at start-up. These two "name" and "dob" are then available for usage in the execution context.

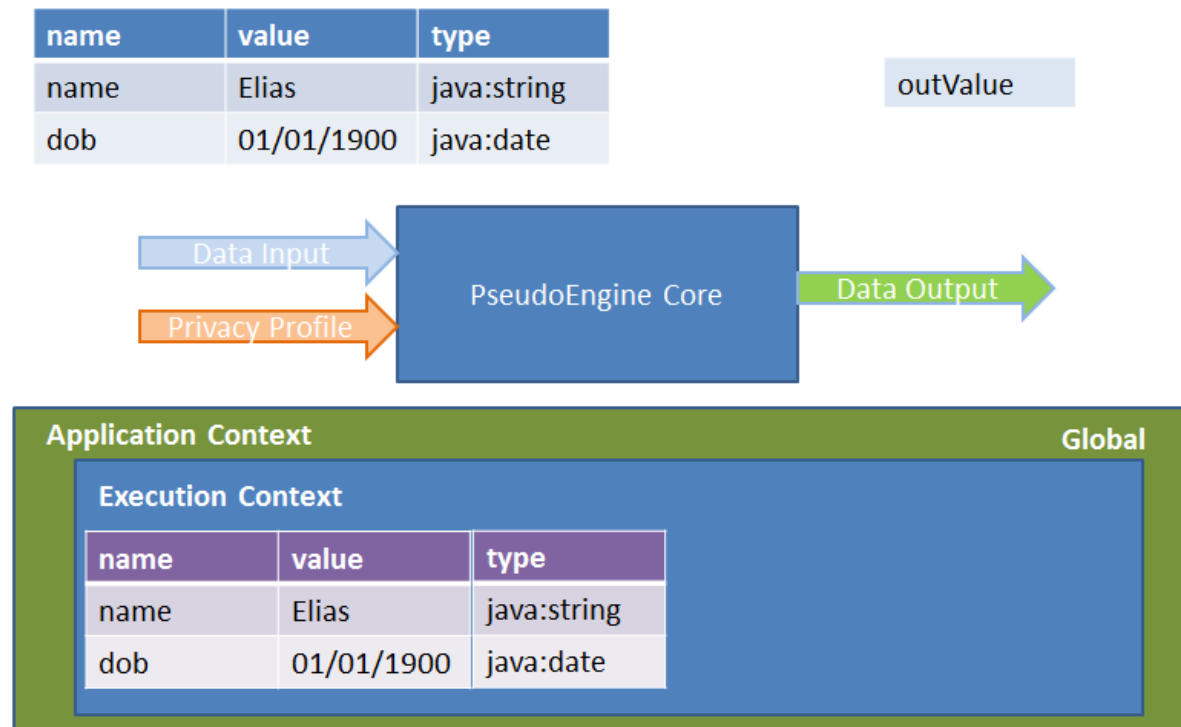


Figure 18 Sample Engine Core Context

The privacy operations in a profile can refer to variables in the context. This allows the use and modification of those variables. The script in Figure 1921 for example will clear the variable “dob” and replace “name” with the string “PSEUDONYMISED”. The resulting context is shown in 22.

```
<process>
  <function ref="core:copy">
    <argument name="source">
      <constant value="PSEUDONYMISED"/>
    </argument>
    <returnValue name="name"/>
  </function>
  <function ref="core:clear">
    <returnValue name="dob"/>
  </function>
</process>
```

Figure 19 Simple pseudonymization example

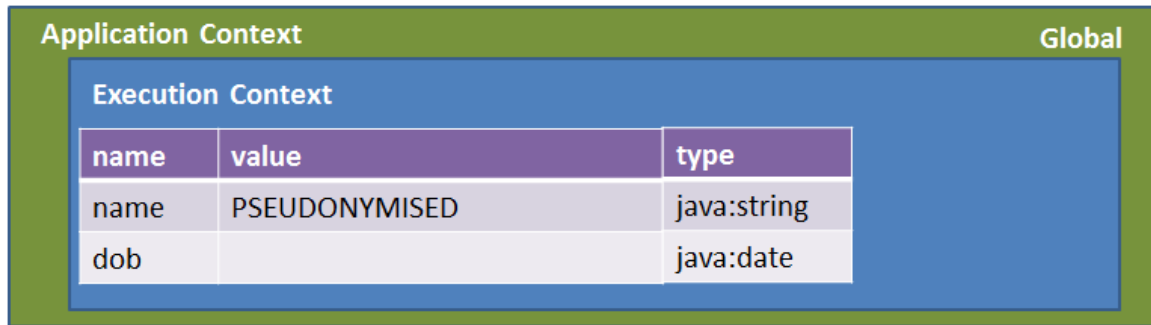


Figure 20 Simple pseudonymization example context

During execution the engine will create further child contexts modelling local variable scope. Each function will for example result in a local context being created. Local contexts can be nested (nested function calls).

In the example from Figure 19 a local function context will be created when the “core:copy” function is called (Figure 20).

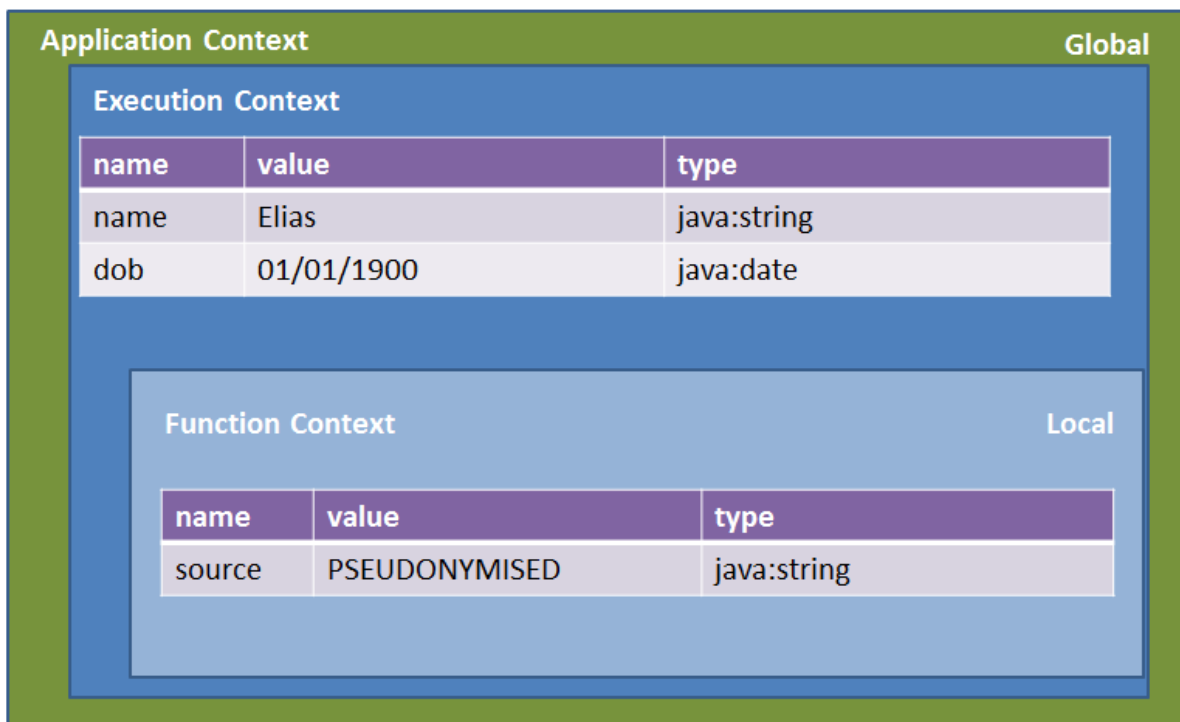


Figure 21 Simple pseudonymization example function context

5.5.2.3 Data mapping and transformation

Through the data mapping extensions the engine is extended with support for stream processing. The mapping in a privacy profile defines how the variables of one row from the data stream are mapped to variables in a local data row context. A row can literally refer to a row (e.g. in CSV stream), but it can also refer to a file, an element in an XML document... The defined transformations will then update these variables and those updated values are then outputted to the output stream.

The data mapping in for example

Figure 4 will result during execution for each data row in context as show in Figure 22.

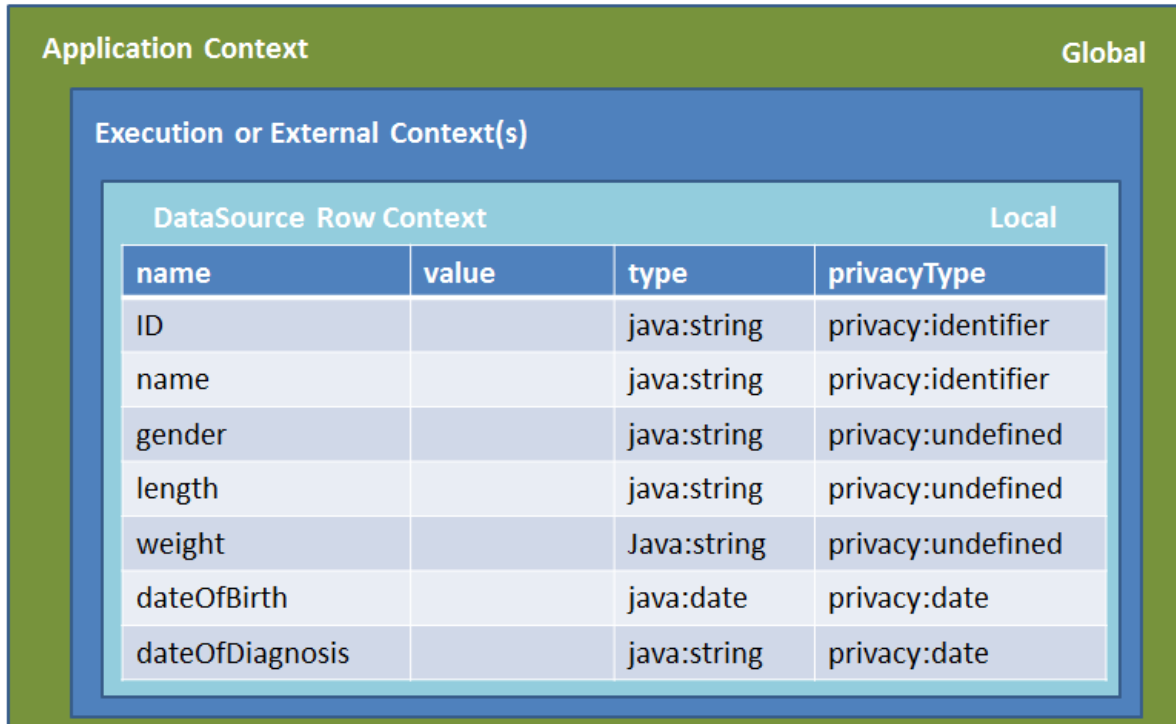


Figure 22 Data Row Context for example CSV Mapping

Through a privacy profile operations are then defined on these row variables.

```
<function ref="core:randomId">
  <argument name="entityIdentifier">
    <variable ref="ID"/>
  </argument>
  <argument name="outputLength">
    <constant type="java:int" value="20" valueType="xsd:string"/>
  </argument>
  <argument name="database">
    <variable ref="projectDatabase"/>
  </argument>
  <argument name="tableName">
    <constant value="patientIDLinkTable"/>
  </argument>
  <argument name="reusePreviouslyGeneratedAnons">
    <constant value="true" type="java:boolean"/>
  </argument>
  <returnValue name="ID"/>
</function>
<forEachVariable itemName="identifier" privacyType="privacy:identifier" condition=
"identifier.updated=='false'">
  <function ref="core:clear">
    <returnValue name="identifier" />
  </function>
</forEachVariable>
```

Figure 23 Privacy Operations

Figure 23 for example defines that the 'ID' field should be replaced by a random identifier and all other identifier fields cleared.

6. Intellectual Property Aspects

In this chapter, we consider some aspects of Intellectual Property Rights (IPR or IP rights), which shall serve as guidelines for transition of the project into the exploitation stage. In part 6.1, we identify exploitable software outcomes of CHIC and suggest suitable licensing options, both for the CHIC components, repository architectures, models and hyper-models, but also for CHIC integrative platform as a whole. In part 6.2, we describe the schemas of managing IPR in the exploitable outcomes of CHIC and focus, in particular, on the licensing issues for hyper-models. In part 6.3, we analyse the protectability of the CHIC repositories by sui generis database rights and the scope of such protection. Part 6.4 describes alternative options of protecting the clinical data in CHIC by IP rights and contractual mechanisms and tackles data ownership issues.

6.1 *CHIC exploitable outputs und software licensing*

6.1.1 CHIC exploitable outcomes

CHIC proposes the development of clinical trial driven tools, services and secure infrastructure that will support the creation of multiscale cancer hyper-models (integrative models)⁴⁶. Consequently, most of the exploitable foreground IP, which was generated under the project, constitutes software components, computer models (or software, implementing the cancer models), visualisation software, data repositories and technological architecture.

The exploitable CHIC outputs, as have been defined by the project partners, and the exploitation plans are described in detail in Deliverable D12.4 – Draft Plan for the Use and Dissemination of Foreground, Section 3 Exploitable CHIC outputs. In total, 22 exploitable outputs have been identified:

- 14 technological/software components,
- 7 models (including both hypo and hyper-models), and
- the CHIC platform as a whole.

Identifying appropriate licensing solutions for the CHIC models and software outputs constitutes one of the most important issues, which deserves particular attention. On the one hand, the licensing model should be adequate to protect the intellectual investment, deployed into creation of the CHIC components, and suit the exploitation plans and interests of the developing parties. On the other hand, CHIC also makes use and to much extent benefits from the use of achievements of other research projects, software and products, placed by the developers into the public domain. Hence, some regard should also be given in direction of contributing the CHIC results to the benefit of e-health research and VPH-research community. Pursuant to the contractual framework of CHIC, and the CHIC Consortium Agreement (CA)⁴⁷ and CHIC Memorandum of Understanding (MoU)⁴⁸, in particular, it is the exclusive rights of the party or parties, who developed the CHIC foreground, to decide on exploitation of its/their foreground and choosing the license model.

At the same time, as also foreseen by the CHIC CA, Article 9.8.6, the project makes wide use of and builds its components on the open source software (OSS). OSS, which in terms of the CHIC CA is licensed under “Controlled License Terms”, may produce licensing implications and limit the exploitation options for the CHIC software outputs. Along with that, use of OSS dependencies and distribution of CHIC components, packaged with such dependencies, in breach of the OSS license terms, combining incompatible libraries or codes

⁴⁶ <http://chic-vph.eu/project/objectives/>

⁴⁷ Articles 8.1, 8.2 CA.

⁴⁸ Article 3.2 MoU.

into one software product, releasing the software product on proprietary basis, when the dependency license allows non-commercial distribution only, may produce legal consequences comprising termination of rights on the use of such dependency, cease and desist actions, copyright litigations⁴⁹. It must also be borne in mind, that upstream licensing (licenses of the software dependencies) may also have licensing implications on the downstream licensing of CHIC components (requiring release of CHIC components under certain license terms), thus leaving developers with no license choices except licensing the component under the license terms of the dependency.

6.1.2 Software licensing issues

In order to support the developing parties with making their license choices and to mitigate the potential license incompatibility risks, the software components and models in CHIC were analysed on subject of license incompatibility issues and advising appropriate license options for downstream licensing. In the result of the legal analysis, the CHIC software components and models, as provided by the developing parties, were identified, the license incompatibility issues were checked for each component, it was explored whether model licenses and combining models into hyper-models may produce licensing implications for the hyper-models, the legal nature of CHIC platform and suggested license solution was identified.

Please see the complete legal analysis “LUH Report on open source license check and software licensing in CHIC” (CHIC software licensing report) in Appendix 6.

The legal analysis produced the following results:

- altogether 23 CHIC software components, including software, repositories, models, hyper-models and CHIC platform were identified (total 23);
- 3 repositories: Clinical Data Repository/UBERN (item 4), Model/ Tool Repository/ICCS (item 9), In Silico Trial Repository/ICCS (item 10) and one data management system: ObTIMA /USAAR (item 1). ObTIMA was added to the list of exploitable outputs already provided in D12.4.10 models and hypermodels;
- 8 software components;
- 1 CHIC platform (item 23)..

The legal analysis reflects only components, which the project parties provided upon the LUH request (email “Software Licensing” of 20 June 2016 with reminders in July). The components/models, the data on which was not delivered by the parties, may be missing and not reflected in the analysis.

6.1.3 Licensing solutions for CHIC software components

In the result, 8 software components were reported: Robust Interactive Multi-label (software), USAAR (item 2), Automatic brain tumor segmentation with a fast Mumford-Shah algorithm (software), USAAR (item 3), Bratumia (software), UBERN (item 5), CRAF (software), FORTH (item 6), Security software component, CUSTODIX (item 13), VPH-HF, CINECA/USFD (item 14), Preprocessing Tool (FORTH) (item 20), Hypermodelling Editor, FORTH (item 22).

As was reported by the parties and may be seen from the CHIC software licensing report (Appendix 6), most of the component come with dependencies released under permissive copyright licenses, like Apache License Version 2, MIT License, BSD 3-Clause License, or weak copyleft licenses, like GNU Lesser General Public License (LGPL), Eclipse Public License (EPL), Common Development and Distribution License (CDDL), Common Public License (CPL), Mozilla Public License. The license incompatibility issues between the dependencies were not identified. The potential license incompatibility issues by downstream

⁴⁹ A. M. St. Laurent, “Understanding Open Source and Free Software Licensing”, O’Reilly, 1 Edition, 2004.

licensing were predicted and license options, advised for the prevention of the legal issues resulting from the use of incompatible codes, were identified.

The components normally use such external dependencies and software libraries by calling the object code (dynamic and static linking), via procedure calls or command line arguments. Use of these dependencies under these licenses by this mechanism of communication does not produce licensing implications for the CHIC components. These licenses do not consider linking as creating a derivative work and allow the originally licensed code to be linked and *“distributed as part of a larger proprietary piece of software, which would not generally be possible under the terms of stronger copyleft licenses.”*⁵⁰ Also, *“command-line arguments are communication mechanisms normally used between two separate programs. So when they are used for communication, the modules normally are separate programs”*⁵¹.

Against these observations, the licensing of the components, which make use of external dependencies or software libraries under these licenses by calling the object code or command line arguments is not restricted and licenses of the dependencies are not binding for these CHIC components. Such CHIC components may be licensed royalty free open source and as executables on proprietary basis. The precondition for this is that if CHIC components are distributed packaged with the dependencies, the license terms for distribution of dependencies codes must be observed.

As open source licenses, these licenses permit redistribution and use of originally licensed codes in source and binary forms, with or without modification, require however, that license terms for distribution of the originally licensed terms be met⁵². Therefore, provided the software components are to be distributed packaged with their software dependencies, the developers would normally be required to keep the license, copyright notices and disclaimers of warranties in the files of dependencies intact and provide license texts along with the distribution (BSD 3-Clause License, MIT License, Apache v2). For the dependencies licensed under EPL, CDDL, CPL, MPL, which require making the originally licensed source code available, the developer, if distributing his component as an executable, would be required to credit the use of such dependencies and their licenses and provide instructions how to get the dependencies source codes. The license, copyright notices and disclaimers of warranties in the files of dependencies must be kept intact.

For the case that developing parties would like to license their components “open source”, the suggested licenses for such components are permissive copyright license Apache License Version 2.0⁵³ (provided licensing under Apache v2 does not cause incompatibility issues, as we identify for some components, e.g. Robust Interactive Multi-label) and weak copyleft license LGPL Version 3. For further details please see CHIC software licensing report (Appendix 6). Observations on suggested open source licenses Apache v2 and LGPL we provide below.

6.1.4 Licensing solutions for CHIC models

As defined in Deliverable No. 7.1 Hypermodelling specifications⁵⁴, 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), it computes certain quantities (data outputs) on the basis of a set of initial quantities (data inputs) and a set of execution logs (control outputs).”* In view of this definition and implementation of cancer model by computer software, the term “model”, as used in this deliverable shall mean software, implementing a particular cancer model. For instance, UOXF, when providing the data for the

⁵⁰ <https://www.mozilla.org/en-US/MPL/2.0/FAQ/>

⁵¹ <https://www.gnu.org/licenses/gpl-faq#MereAggregation>

⁵² See BSD 3-Clause License, Section 2 Apache 2.0, MIT License, Section 2.1 MPL 2.0.

⁵³ <https://opensource.org/licenses/Apache-2.0>

⁵⁴ http://chic-vph.eu/uploads/media/D7-1_Hypermodelling_specifications.pdf

Vasculature models of Nephroblastoma and Vasculature model of Lung cancer, described the substance of the model as “Software implementing the vasculature hypomodel for nephroblastoma” and “software implementing the vasculature hypomodel for lung cancer”, respectively⁵⁵.

In the result of the legal analysis, 8 models and hyper-models were identified: Lung cancer oncosimulator (mechanistic)(model), ICCS (item 7), Glioblastoma oncosimulator (model), ICCS (item 8), NSCLC machine learning based response predictor (model), ICCS (item 11), Nephroblastoma oncosimulator (model), ICCS (item 12), Vasculature Nephroblastoma Software implementing the vasculature hypomodel for nephroblastoma, UOXF (item 15), Vasculature Lung software implementing the vasculature hypomodel for lung cancer, UOXF (item 16), Universal growth and response to treatment, UNITO (item 17), Molecular Network Model, UPENN (item 18), Biomechanical simulator (UBERN) (item 19), Metabolic Model, FORTH (item 21).

Nine of these 10 CHIC models were built with the use of external codes or software libraries, licensed under permissive copyright licenses, like Apache v2, BSD 3-Clause, MIT License. Some models rely on MUSCLE 2, licensed under weak copyleft license LGPL Version 3⁵⁶. Normally, the models use their dependencies by calling the object code. As noted above, most OSS licenses do not consider linking as creating a derivative or modified work, bound by the dependency license; thus use of dependencies under permissive copyright licenses does not produce licensing implications for the linked code. For instance, in terms of Apache v2 *“For the purposes of this License, Derivative Works shall not include works that remain separable from, or merely link (or bind by name) to the interfaces of, the Work and Derivative Works thereof.”*⁵⁷

Also linking to LGPL covered MUSCLE does not affect licensing of the CHIC models, because LGPL v3 allows a developer to create a combined work *“produced by combining or linking an Application with the Library”* and to *“convey a Combined Work under terms of your choice that, taken together, effectively do not restrict modification of the portions of the Library contained in the Combined Work and reverse engineering for debugging such modifications, if...”*⁵⁸ a developer credits the use of LGPL library and provides a possibility how to get the source code for the library, as Section 4 LGPL provides.

Out of the models, analysed so far, only Biomechanical simulator, UBERN (item 19) makes use of the dependencies, some of which are licensed under strong copyleft licenses GPL v2 and GPL v3. The Biomechanical simulator communicates with the GPL licensed codes by calling the object codes. In opinion of the FSF, *“Linking a GPL covered work statically or dynamically with other modules is making a combined work based on the GPL covered work. Thus, the terms and conditions of the GNU General Public License cover the whole combination”*⁵⁹. Following this position, distribution of Biomechanical simulator packaged with its GPL dependencies would require licensing of the simulator under GPL. Since GPL v2, applicable to CodeSynthesis XSD⁶⁰, one of the simulator dependencies, allows upgrade to a later version⁶¹, GPL v3, which is considered more flexible and compatible with the other open source licenses is suggested for this model.

⁵⁵ Items 16,17, LUH Software licensing report, Appendix 2.

⁵⁶ <http://apps.man.poznan.pl/trac/muscle>

⁵⁷ Apache License Version 2.0, available at: <https://opensource.org/licenses/Apache-2.0>

⁵⁸ Section 4 LGPL Version 3, available at: <https://opensource.org/licenses/LGPL-3.0>

⁵⁹ <https://www.gnu.org/licenses/gpl-faq#GPLStaticVsDynamic>

⁶⁰ <http://www.codesynthesis.com/products/xsd/>

⁶¹ Section 9 GPL v2, available at: <https://opensource.org/licenses/GPL-2.0>

6.1.5 Licensing solutions for CHIC hyper-models

In the context of CHIC, a hyper-model (or composite model or integrative model) is defined as *“a model that emerges from the composition and orchestration of multiple hypomodels, each one of which is capable of simulating a specific entity or phenomenon. The hypermodel can simulate an entity or phenomenon that may be more complex than the ones simulated by each separate simpler model.”*⁶²

As indicated by the modelling parties, the communication between models in hyper-models (or integration of models into hyper-models) occurs mostly via MUSCLE API. For example, the Lung cancer oncosimulator (mechanistic) (model), developed by ICCS, which is part of the Lung cancer multimodeler hypermodel, communicates with (a) the molecular model (UPENN): static/sequential communication through command line argument; (b) the preprocessing tool (FORTH): static/sequential communication through command line argument; (c) the Biomechanics simulator (UBERN): dynamic/iterative communication through MUSCLE library; (d) the vasculature model (OXFORD): dynamic/iterative communication through MUSCLE library; and (e) the metabolic model(FORTH): dynamic/iterative communication through MUSCLE library⁶³. As stated: *“...pipes, sockets and command-line arguments are communication mechanisms normally used between two separate programs. So when they are used for communication, the modules normally are separate programs.”*⁶⁴

Based on this interpretation, when models communicate with each other via command line arguments, the models are considered as separate programs and license of one model should not produce and affect licensing of other models and a resulting hyper-model as a whole.

Another mechanism of interaction, used for combining models in CHIC, as communication via MUSCLE library. Also this method of communication does not create licensing implications. MUSCLE LGPLv3 license, Section 4, allows *“combining or linking an Application with the Library”* and to *“convey a Combined Work under terms of your choice, that, taken together, effectively do not restrict modification of the portions of the Library contained in the Combined Work and reverse engineering for debugging such modifications”*⁶⁵

Against this technical background and licensing considerations, licenses of the models combined into a hyper-model should not produce licensing implications on licensing of a hyper-model. A hyper-model, combined from models communicating via MUSCLE library, may go under its own license. However, as Section 4 LGPL v3 requires, the hyper-model license must allow modification of the MUSCLE library contained in the hyper-model and reverse engineering for debugging such modifications.

6.1.6 Licensing solutions for the CHIC repository architectures

Within the CHIC software components, 3 were reported a repositories architectures (Clinical Data Repository/UBERN (item 4), Model/ Tool Repository/ICCS (item 9), In Silico Trial Repository/ICCS) and one, namely ObTIMA, USAAR (item 1) as a data management system.

Construction of repositories architectures relies on the standard methods of software development and legal analysis follows the same legal rules, as for software components. The repositories are usually built using a large number of dependencies. The ICCS In Silico

⁶² Section 3.4.3 CHIC Deliverable No. 7.1 Hypermodelling specifications, available at: http://chic-vph.eu/uploads/media/D7-1_Hypermodelling_specifications.pdf

⁶³ Item 7, LUH Software licensing report, Appendix 6.

⁶⁴ <https://www.gnu.org/licenses/gpl-faq#MereAggregation>

⁶⁵ <https://opensource.org/licenses/LGPL-3.0>

Trial repository and Model/Tool repository are built upon MySQL database architecture. In case of USAAR ObTiMA, the dependencies are licensed, either under permissive copyright licenses, i.e. Apache v2, BSD, MIT license, or weak copyleft licenses, i.e. EPL, CDDL, CPL, LGPL. ObTiMA links to the dependencies by calling the object code. In terms of these licenses, linking or calling the object code, does not produce licensing implications for the software, which links against them. Therefore, there is no binding license for ObTiMA because of the dependencies licenses.

Suggested open source licenses for ObTiMA are permissive copyright licenses: BSD 3-Clause License, MIT License. Licensing under GPL would create incompatibility problems with EPL, CDDL, CPL⁶⁶. If packaged with LGPL dependencies, licensing under Apache might create incompatibility issues with LGPL, because Apache does not tolerate LGPL in Apache products⁶⁷. If distributed with LGPL dependencies, the license must allow modification and reverse engineering of LGPL libraries for the customer's own use (para 6 LGPLv.2.1). Distribution in object code (on a commercial basis) and source code is possible, but where distributed packaged with dependencies, copyright license notices and disclaimers of warranties in files of the dependencies must be preserved. For the codes under EPL, CDDL, LGPL, CPL instructions should be given on how to obtain the source code.

Licensing of the UBERN Clinical Data Repository (CDR) (item 4) is also not restricted by the dependencies. The CDR makes use of dependencies under permissive copyright licenses (Apache v2, MIT, BSD) by calling the object code or via http calls. Use of the dependencies via these modes of communication does not produce licensing implications for CDR. Distribution in object code on a proprietary basis and source code is possible. A suggested open source license is Apache v2 .

The ICCS Model/ Tool Repository and In Silico Trial Repository are built on MySQL database architecture. *"MySQL is the world's most popular open source database"*⁶⁸. It is licensed under GPLv2. Although, GPL v2 is a strong copyleft license, however, My SQL is installed separately and used by the repositories via Unix sockets. In terms of GPL, sockets *"are communication mechanisms normally used between two separate programs. So when they are used for communication, the modules normally are separate programs"*⁶⁹. Based upon this position, use of MySQL does not affect licensing of the repositories. The other dependencies, used by ICCS repositories, are released under permissive copyright licenses (BSD, MIT, Apache) and using them by calling the object codes, or command line arguments does not produce licensing implications for the ICCS repositories.

The repositories may be distributed in object code on proprietary basis and open source. Suggested open source licenses are LGPL v3, or Apache v2.

In respect to the repository architectures, it must be noted that protection and licensing repositories software architectures as copyright works may occur in parallel and does not interfere with potential protection of the repository contents by sui generis database rights, as we consider below.

6.1.7 Suggested open source licenses

It has become common practice in various research projects, where the aim is to increase the project sustainability and dissemination of project results, to release part or all software components open source⁷⁰. Should the CHIC Parties follow this example and be willing to

⁶⁶ <https://www.gnu.org/licenses/license-list.en.html>

⁶⁷ <http://www.apache.org/legal/resolved.html>

⁶⁸ <https://www.mysql.com/products/>

⁶⁹ <https://www.gnu.org/licenses/gpl-faq#MereAggregation>

⁷⁰ Linked2Safety, FP7-288328, <https://github.com/linked2safety/code>; MyHealthAvatar, GA N 600929, RESTful API for retrieving clinical data in the MyHealthAvatar platform: <https://github.com/sgsfak/mha-clinica-api>.

release their codes “open source”, the open source licenses: Apache License Version 2.0⁷¹ and the GNU Lesser General Public License, version 3.0, LGPL v3⁷², and General Public License, version 3.0, GPL v 3⁷³ are suggested.

6.1.7.1 Apache License, Version 2.0

An Apache license would make a suitable option for those components, which the developers intend to release open source and do not exclude the use of their codes in proprietary products.

Apache License v2 is a classic open source license; it allows licensing software products both “open source” and on a proprietary basis, and does not require the source code to be disclosed. The fairly relaxed license terms make it flexible and compatible with many forms of open source licenses⁷⁴. Apache v2 fits well for communications-oriented software, such as used in the Apache HTTP server and adheres to standards such as HTTP protocol.

Components, once released under Apache, may be distributed as binary executables on a proprietary basis and/or “open source” (the terms for distributing the Apache code must be observed). Fees for royalties and support may be charged⁷⁵. The provider of warranty protection needs to be aware that he accepts the responsibility for the whole software product, including modules written by the other parties.

It is an essential feature of Apache that it allows commercialization of the developed software. It makes licensing under Apache v2 suitable for research projects in order to explore the technology, where, upon successful implementations of research ideas commercialization is envisaged.

On the other hand, licensing under Apache may contain a risk, that any third party may convert CHIC components into proprietary products or create modified versions and distribute such modified versions on a proprietary basis. The Apache license allows this. The only condition is that the code, originally released under Apache, must remain under Apache throughout the whole subsequent distribution. Also the requirement of the Apache License that the names of the licensors may not be used for endorsement purposes of derivative products provides some level of protection⁷⁶.

In order to prevent conversion of the codes developed in CHIC into proprietary products and to keep the codes “open” for the research community, licenses with weak copyleft, such as LGPL, or strong copyleft, such as GPL, are recommended as more suitable options. These licenses will help to keep the codes originally released “open source” as “open source” in all subsequent distributions of verbatim copies and modified versions.

6.1.7.2 GNU Lesser General Public License (LGPL), Version 3.0

The GNU Lesser General Public License, version 3.0 (LGPL v3) is a weak copyleft license. It allows linking LGPL codes with proprietary modules and distributing combined software on a proprietary basis, but provides that a LGPL code remains under LGPL and stays accessible in its source form.

LGPL v3 is the latest version of the LGPL. It is a free software license with weak copyleft. It permits linking with nonfree modules and the creation of proprietary software⁷⁷. LGPL v3 license allows “*combining or linking an Application with the Library*” and to “*convey a*

⁷¹ <https://opensource.org/licenses/Apache-2.0>

⁷² <https://opensource.org/licenses/LGPL-3.0>

⁷³ <https://www.gnu.org/licenses/gpl-3.0.en.html>

⁷⁴ St Laurent, Andrew M., Understanding Open Source and Free Software Licensing, 2004, p. 32 et seq.

⁷⁵ Section 9 Apache v2.

⁷⁶ St Laurent, Andrew M., Understanding Open Source and Free Software Licensing, 2004, p. 32

⁷⁷ <https://www.gnu.org/licenses/license-list.en.html>

*Combined Work under terms of your choice, that, taken together, effectively do not restrict modification of the portions of the Library contained in the Combined Work and reverse engineering for debugging such modifications*⁷⁸ What is required for distribution, is that the use of LGPL licensed code and LGPL license are credited and that a possibility to get the LGPL licensed code is provided to the users⁷⁹.

In terms of exploiting CHIC software components licensed under LGPL, this means that such components may potentially be combined into and be distributed as part of proprietary products. However, the LGPL licensed codes will stay under LGPL and must be accessible to the users, irrespective of the mode of distribution (as an executable on a proprietary basis or “open source”). If released under LGPL, only royalty-free distribution is allowed. Warranties and support may be provided for a fee, but at the sole responsibility of the party, offering such protection.

6.1.7.3 GNU General Public License (GPL), Version 3

GNU General Public License (GPL), Version 3 (GPL v3), is suggested as an optimal license solution for those parties, who intend to license their software developments in a way to keep the codes open for subsequent development and wish to exclude the use and licensing their codes in proprietary products.

GPL v3 is the latest version of the GNU GPL: a free software license and a copyleft license⁸⁰. Due to its strong copyleft, GPL v3 is designed to make sure that developers have the freedom to distribute copies of GPL programs (and may charge for them), that they receive source code or can get it, that they can change the software or use pieces of it in new programs, and that they are aware of these rights⁸¹.

The components, released under GPL, may be distributed on a royalty-free basis only; distribution in object code would require making the source code accessible. Fees for physical distribution (download) of copies and fees for offering warranties and protection are allowed. A party, offering warranty protection for the software, which was developed with the use of open source modules, provided by third parties, acts at its own risk and assumes responsibility (and enters into a duty to fix bugs) for the whole software package (including third party modules).

6.1.8 Legal nature and licensing solution for the CHIC platform

The CHIC components, including repositories, hypermodelling environment, visualization and modelling toolkits, security framework are deployed in the CHIC architecture in the private cloud hosted by FORTH⁸². The communication and interaction among the components occurs via http calls, exchange of files (output of one model serves as an input into another), command line arguments, database protocols, etc. *“Pipes, sockets and command-line arguments are communication mechanisms normally used between two separate programs. So when they are used for communication, the modules normally are separate programs.”⁸³* Hence, by these modes of communication, even if licenses of some individual CHIC components are incompatible with other components, the license terms of one component do not produce licensing implications for those others.

Considering that CHIC components interact as separate and individual programs, but are integrated on the CHIC platform, the integrated CHIC platform may be considered as an “aggregate”. An “aggregate” is composed “of a number of separate programs, distributed

⁷⁸ <https://opensource.org/licenses/LGPL-3.0>

⁷⁹ Section 4 LGPL v3.

⁸⁰ <https://www.gnu.org/licenses/license-list.en.html>

⁸¹ Preamble to GPL v3: <https://www.gnu.org/licenses/gpl-3.0.en.html>.

⁸² CHIC Deliverable No. 10.1, The CHIC Portal, Section 5.2 Architecture – Components, p. 18.

⁸³ <https://www.gnu.org/licenses/gpl-faq#MereAggregation>

together on the same CD-ROM or other media.”⁸⁴ Pursuant to criterion 1 of Open Source Definition⁸⁵, an open source license shall “*not restrict any party from selling or giving away the software as a component of an aggregate software distribution containing programs from several different sources.*” Accordingly, also the GPL permits creating and distributing an aggregate, even when the licenses of software components are non-free or GPL-incompatible⁸⁶. The one condition, which must be considered when licensing an “aggregate”, is that license of the aggregate shall not prohibit or restrict the user “*from exercising rights that each program’s individual license would grant them.*”⁸⁷ To satisfy this condition, the CHIC platform shall be licensed as an aggregate in such a way that individual components go under their own licenses and the users have a possibility to use individual components within the scope permitted by each component license.

6.2 Managing IPR in CHIC

6.2.1 CHIC Memorandum of Understanding

One of the basic instruments, which was put in place with regard to managing the IPR in the CHIC project, is the Memorandum of Understanding on Disposal of Intellectual Property Rights in CHIC (IPR MoU). The was indicated as desirable by the Project Reviewers following the CHIC interim review⁸⁸ and supplements the default rules of the FP7 EC-GA (Annex II) and provisions of the CHIC Consortium Agreement (CA) in clarifying the governance of IPR in CHIC. The main innovation was the introduction of the legal regime of composite ownership (which is of particular importance for the hyper-models and other integrative works) and reaching agreement among the parties on the rules for exercise of IPR in the works of composite ownership. Other matters dealt with were legal issues associated with the grant of Access Rights, and in particular, Access Rights to software. The grant of the Access Rights to the Foreground for Use has been made subject to conclusion of a written agreement, a typical licensing practice providing for evidential certainty in the project. Further legal issues relating to the use and grant of Access Rights to software are also addressed, including the use of open source software, compliance with license terms and license incompatibility issues.

The draft IPR MoU originated in the result of legal analysis of CHIC IPR issues conducted in deliverable D.4.3.1 “Development of the data protection and copyright framework for CHIC first iteration” (M14)⁸⁹. The first draft was attached to D.4.3.1 as Appendix 5. The IPR MoU was agreed by the Parties in Version 3.0 as of January 2015 and entered into force with last signature made on 30 June 2015. A copy was provided to the Commission at the third CHIC annual review in July 2015.

6.2.2 Managing IPR in hyper-models

The ownership issues and exercise of IPR in hyper-models fall under the rules of IPR MoU. As follows from the analysed materials⁹⁰, insofar the hyper-models for the two tumour types: lung cancer and Nephroblastoma have been generated in CHIC. The multi-modeller lung cancer (and also Nephroblastoma) hyper-model is composed of: (a) mechanistic model of ICCS; (b) vasculature model of UOXF; (c) biomechanical model of UBERN; (d) molecular model of UPENN and (e) metabolic model of FORTH.

⁸⁴ Ibid.

⁸⁵ <https://opensource.org/osd>

⁸⁶ <https://www.gnu.org/licenses/gpl-faq#MereAggregation>

⁸⁷ Ibid.

⁸⁸ Period 1 Review (November 2013): Reviewers’ Consolidated Report, p. 3.

⁸⁹ http://chic-vph.eu/uploads/media/D4-3-1_Development_of_the_data_protection.pdf

⁹⁰ LUH Software Licensing Report.

According to Clause 3.2 CHIC IPR MoU, the contributing parties: ICCS, FORTH, UPENN, UBERN and UOXF own the IPR, applicable to the hyper-models in question. In principle, exploitation and licensing of these hyper-models requires agreement of all the contributing parties and it is a privilege of these parties to agree on the exploitation and licensing strategy for the hyper-models.

However, shall the modelling parties fail to enter into such agreement, then in accordance with paragraph 5 Clause 3.2 CHIC IPR MoU, the default rule, stipulated by Article 8.2 CA for the exploitation of joint foreground shall apply. It means, each of the contributing parties (ICCS, FORTH, UPENN, UBERN and UOXF) shall be entitled to Use the hyper-models it sees fit, *“and to grant non-exclusive licenses, without obtaining any consent from, paying compensation to, or otherwise accounting to”* each other. It means, in the absence of agreement between the modelling parties, each party (ICCS, FORTH, UPENN, UBERN and UOXF) will be entitled to exploit and license the hyper-models on its own, provided such individual exploitation is possible beyond the technical framework of CHIC.

However, with regard to the potential exploitation of hyper-models it must be borne in mind that hyper-models as such or as part of CHIC hyper-modelling environment, if intended for use in clinical practice would fall under the medical device regime⁹¹ and before putting into exploitation will need to be certified. Before the validation of hyper-models for the clinical use and such certification is realistic, the potential exploitation of hyper-models in research and improvement the technology make more plausible options.

6.2.3 Access Rights to hyper-models

There is an express interest on part of the clinical parties to explore the hyper-models in the clinical practice. At this stage, such use is intended for the purpose of validating and fine-tuning the technology, and is not intended for clinical decision making. For the use of hyper-models as clinical decision support tools the medical device regime would apply⁹².

The CHIC parties have an option to obtain the rights on use of the hyper-models beyond the domain of CHIC under the umbrella of Access Rights. The Access Rights for Use are governed by Section 9.4 CHIC CA. According to Article 9.4.1 CA, Access Rights to Foreground for internal research activities and Access Rights to Foreground (which is not patented) for Use shall be granted on royalty-free basis. Pursuant to Article 9.4.3 a request for Access Rights may be made up to 12 months after the end of the project. The grant of Access Rights must be evidenced in writing (Clause 3.4 MoU).

Should the clinical parties of CHIC, i.e. USAAR, KU Leuven, UNITO, wish to explore the hyper-models in the clinical setting after the project ends, they shall submit a written request for Access Rights for Use in the form, as provided in Attachment 1 CHIC IPR MoU. In particular, the request shall identify the hyper-model(s) in question, reasons why Access Rights are needed, the scope of rights and scope of use (territory, duration), the right to sublicense, if necessary. The request must be addressed to the modeling parties, who contribute and own rights in the hyper-models. Once the Access Rights are granted, the modeling parties signify the grant of Access Rights by signing and returning signed request to the requesting party. Such exchange of the request for Access Rights with signatures of

⁹¹ Directive 93/42/EEC of 14 June 1993 concerning medical devices, OJ L 169, 12.7.1993, p. 1, consolidated version; EC, Medical Devices Guidance Document, Qualification and Classification of stand alone software, MEDDEV 2.1/6, 2012, January, <ec.europa.eu/health/medical-devices/files/meddev/2_1_6_ol_en.pdf> 29.06.2016.

⁹² *The legal assessment of CHIC models under the medical device regime has been done with the outcome that cancer models released as stand alone software are most likely to qualify as medical devices on their own. Also, the CHIC system, including hyper-modelling framework, data repositories, models, image processing tools, if released into use is to qualify as decision support software. See: I. Lishchuk, M. Stauch “Cancer models as medical support tools”, to be published in Conference Proceedings, Herbst Akademie, Hamburg, 14-17 September 2016.*

the requesting and granting parties constitutes conclusion of a license agreement. The time frame for the request of Access Rights is limited to 12 months after the project ends.

Since deployment and running hyper-models in clinical setting requires the software platform CRAF (Clinical Research Application Framework)⁹³, a request for Access Rights to hyper-models shall be done in conjunction with the request for Access Rights to CRAF, which must be addressed to FORTH.

6.3 Protection of CHIC repositories by sui generis database rights

6.3.1 CHIC repositories

CHIC proposes the development of software tools, services and infrastructures that will support and facilitate the creation of multiscale cancer hypermodels (integrative models).⁹⁴ The simulation of cancer progression is achieved by multiscaling. Multiscaling is realised by combining models, developed by several modelling parties, each of which simulates a biological process at a particular time-space scale, into hyper-models, thus capturing cancer progression across several biological scales at once. The simulations are run against the clinical data, where the output of one model serves as an input into another. The model predictions are then pre-validated during in silico trials. The CHIC repositories constitute an integral and essential part of the CHIC modelling environment, which along with the models and modelling toolkits are required to perform this analysis. The CHIC repositories constitute storage facilities for the clinical data, used for running the simulations, the models, provide by several parties, and host the results of in-silico trials⁹⁵.

The three repositories are integrated into the hyper-modelling framework of CHIC:

- Clinical Data Repository/UBERN (item 4)
- Model/ Tool Repository/ICCS (item 9)
- In Silico Trial Repository/ICCS (item 10)⁹⁶.

These CHIC repositories provide software architecture to store the necessary elements (either clinical data or models or model outputs). The repositories are deployed in the private cloud infrastructure provided by FORTH. This allows much of flexibility in terms of computational power, storage space and networking⁹⁷.

Provided the one or another repository qualifies as a database in the meaning of Database Directive and stands the requirements for protection by sui generis database rights, the contents of such repository may also be protected by sui generis database rights.

6.3.2 Sui generis database rights

The Database Directive offers sui generis protection to databases, which constitute “a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means”⁹⁸ and which shows “that there has been qualitatively and/or quantitatively a substantial investment in either the

⁹³ The Clinical Research Application Framework (CRAF) is a software platform installed at the clinical side that supports the clinicians in using the hypermodels built and deployed in the CHIC environment for clinical research purposes. D. Deliverable No. 12.4, Draft Plan for the Use and Dissemination of Foreground, p.41 et seq.

⁹⁴ <http://www.chic-vph.eu/>

⁹⁵ CHIC Deliverable D8.2–Prototype implementation of the CHIC repositories.

⁹⁶ LUH software licensing report.

⁹⁷ CHIC Deliverable D8.2–Prototype implementation of the CHIC repositories.

⁹⁸ Article 1(2) Database Directive.

*obtaining, verification or presentation of the contents*⁹⁹. Such protection allows the maker of the database “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”¹⁰⁰. The sui generis database rights pass to the maker of the database, namely “the person who takes the initiative and the risk of investing”, excluding subcontractors¹⁰¹.

The databases may contain “works” in the meaning of traditional copyright law, including literary, artistic and other such works and/or “data”, i.e. data understandable to humans, as contrasted to raw data, which is independent of meaning and has not yet reached the status of human intelligible information.¹⁰² In considering the database definition, the Court of Justice of the European Union (CJEU) affirmed the broad scope of the Database Directive’s wording and explained that a database is “any collection of works, data or other materials, separable from one another without the value of their contents being affected, including a method or system of some sort for the retrieval of each of its constituent materials.”¹⁰³ However, the Court placed great emphasis on the requirement that the materials collected are “independent”, i.e. “separable from one another without their informative, literary, artistic, musical or other value being affected.”¹⁰⁴

In the light of these observations, in order to constitute a database, it is sufficient if the data is organized and individually accessible and separable.¹⁰⁵ However, databases are protected not because they qualify as databases per se. Rather protection accrues to the investment, deployed in “either the obtaining, verification or presentation of the contents” and thus making elements of the contents accessible and re-usable individually, which deserves and obtains protection. Such investment can take the form of time, financial resources, personnel, or technical means deployed in making the database¹⁰⁶.

In this regard, though, collecting the data into the database from one source alone would not suffice. The CJEU demands that the investment in obtaining the data must be made to “seek out existing independent materials and collect them in the database”.¹⁰⁷ An investment in “the creation of materials which make up the contents of a database”¹⁰⁸ is deemed insufficient. As a result, databases, constructed out of data created by one entity only, the so-called “single source databases”¹⁰⁹ do not achieve sui generis database protection, unless they prove a substantial investment in the verification or presentation of the contents.

The investment made in preparing and/or presenting the data from clinical trials for research may justify such protection. First, before entering the research environment, the clinical trial data must undergo an extensive verification process. The latter encompasses steps taken to ensure the information is reliable, but such investment excludes verification of information being created.¹¹⁰ Second, data verification is a separate process and follows the stage of

⁹⁹ Article 7 Database Directive.

¹⁰⁰ Ibid.

¹⁰¹ Recital 41 Database Directive.

¹⁰² Estelle Derclaye, *The Legal Protection of Databases*, p. 58 et seq (2008).

¹⁰³ See Case C-444/02 *Fixtures Marketing Ltd v Organismos prognostikon agonon podosfairou AE (OPAP)*, para 32.

¹⁰⁴ Ibid, para 29.

¹⁰⁵ Davison, *The Legal Protection of Databases*, p. 70 et seq.

¹⁰⁶ Recital 39 Database Directive.

¹⁰⁷ Case C-203/02 *The British Horseracing Board Ltd and Others v William Hill Organization Ltd.*, para 42.

¹⁰⁸ Case C-203/02 *The British Horseracing Board Ltd and Others v William Hill Organization Ltd.*, para 42. See also Jasper A. Bovenberg, *Property Rights in Blood, Genes & Data: Naturally Yours?* p. 176 et seq. (2006).

¹⁰⁹ European Commission, DG Internal Market and Services Working Paper – First evaluation of Directive 96/9/EC on the legal protection of databases, 2005, p. 14, available at:

http://ec.europa.eu/internal_market/copyright/docs/databases/evaluation_report_en.pdf.

¹¹⁰ Estelle Derclaye, *The Legal Protection of Databases*, p. 97 (2008).

obtaining/creating the original data. For its part, “*presentation*” means the process of structuring the data and making it accessible to others. The creation of an index of the repository contents on the user interface can fulfil the requirements of an investment in the presentation of the contents.¹¹¹

Finally, the investment must also be of a “*qualitatively and/or quantitatively*” substantial nature.¹¹² The Database Directive does not define “*substantial*” and neither has the CJEU ruled on the matter. However, the Preamble of the Directive indicates that, “*as a rule, the compilation of several recordings of musical performances on a CD (...) does not represent a substantial enough investment to be eligible under the sui generis right*”.¹¹³ Member States generally adopt a low level approach to the requirement, and the Advocate General has taken the same stance.¹¹⁴ As regards the quantitative and/or qualitative qualification, these are understood to mean investments quantifiable and not-quantifiable, respectively, such as money on the one hand and intellectual effort on the other.¹¹⁵

The database right protection extends against unauthorized extraction and re-utilization of a substantial part, evaluated qualitatively and/or quantitatively, of the contents of the database,¹¹⁶ or the repeated and systematic use of insubstantial parts that infringes upon the legitimate interests of the database maker.¹¹⁷ The CJEU has held that a substantial part, evaluated quantitatively, refers to the volume of data in relation to the database as a whole.¹¹⁸ Evaluated qualitatively, a substantial part refers to the scale of investment in obtaining the contents.¹¹⁹

“*Extraction*” is defined as “*the permanent or temporary transfer (...) to another medium by any means or in any form*”.¹²⁰ Extraction does not require physical copying of contents, and neither does it require the intention of creating a competing product.¹²¹ For example, consulting a protected database, assessing its contents individually and then using it for the creation of a new database that expands upon the contents of the first database can constitute extraction. Specifically, the CJEU held that consulting a database of poetry titles in order to create a database of poems can constitute extraction.¹²²

“*Re-utilization*” is defined as “*any form of making available to the public (...) by the distribution of copies, by renting, by on-line or other forms of transmission*”, albeit subject to the principle of exhaustion.¹²³

As regards the above mentioned repeated and systematic extraction/reutilization of insubstantial parts, the CJEU held that such use is prohibited, “*the cumulative effect of which would be to seriously prejudice the investment made by the maker of the database just as the extractions and/or re-utilisations*”¹²⁴ – which is the case where either the database as a

¹¹¹ Ibid, p. 97 et seq..

¹¹² Article 7(1) Database Directive.

¹¹³ Recital 19 Database Directive.

¹¹⁴ Estelle Derclaye, *The Legal Protection of Databases*, pp. 75 et seqq. (2008).

¹¹⁵ Case C-338/02 *Fixtures Marketing Ltd v Svenska Spel AB*, para 28.

¹¹⁶ Article 7(1) Database Directive.

¹¹⁷ Article 7(5) Database Directive.

¹¹⁸ Case C-203/02 *The British Horseracing Board Ltd and Others v William Hill Organization Ltd.*, para 70.

¹¹⁹ Case C-203/02 *The British Horseracing Board Ltd and Others v William Hill Organization Ltd.*, para 71.

¹²⁰ Article 7(2)(a) Database Directive.

¹²¹ Diane Rowland, Uta Kohl and Andrew Charlesworth, *Information Technology Law*, p. 390 et seq. (4th ed, 2012).

¹²² Case C-304/07 *Directmedia Publishing GmbH v Albert-Ludwigs-Universität Freiburg*, paras 9 et seqq.

¹²³ Article 7(2)(b) Database Directive.

¹²⁴ Case C-203/02 *The British Horseracing Board Ltd and Others v William Hill Organization Ltd.*, para 86.

whole or in substantial parts is basically reconstituted by the systematic and repeated efforts of the user.¹²⁵

6.3.3 Applicability of sui generis database rights to the CHIC repositories

6.3.3.1 Clinical Data Repository

Protection by the sui generis database right can be considered as a plausible option for protecting the Clinical Data Repository (CDR), created by UBERN. To enjoy such protection, the CDR must qualify as a “database” and show a significant investment in either “*the obtaining, verification or presentation*” of its contents.

The CHIC Clinical Data Repository stores the clinical data, which after the necessary de-identification enters the domain of CHIC. The CHIC data repository hosts data categorized per data type.

Table 1: Data types, standards, sources

Data type	Standard	Source
clinical data (pathological and also outcome)	CDISC, ODM	ObTiMA (ontology--- based Clinical trial management system)
imaging data (post-processed, segmented, etc.)	DICOM, Metalmage, Nifti, Analyze	Local PACS (Picture Archiving And Communication System)
Genetic Molecular data	SOFT, MINiML	Platform specific (e.g. Affymetrix) files as generated by the appropriate equipment.

Copied from CHIC Deliverable No. 8.2 Prototype implementation of the CHIC repositories¹²⁶

The datasets for each type are accessible individually so that the data corresponding to the model parameters may be chosen. Indeed, the datasets must be stored and be accessible and re-usable separately and individually, because each particular dataset must be in the form capable of being used as an input into a model. The CDR, storing the datasets arranged in a way and individually accessible by electronic means shall fit into the definition of a database.

The fact that the datasets are collected from different parties and come from different sources, the data is validated and verified, the repository is built from the experience of constructing data repositories in other medical projects¹²⁷ should be sufficient to prove the requisite investment in “*either the obtaining, verification or presentation*” of its contents.

Whereas the data is provided by the clinical parties, who also make the necessary verification, the sui generis protection may apply to the CDR in CHIC in reward for the investment made in presenting the repository contents. As noted, presenting the data means structuring and making it accessible. The data in the CDR is arranged as per the data types, as we considered above. The upload and versioning, linking, annotation, search, validation, data organisation, download are the services, that make the data of the CDR accessible for

¹²⁵ Case C-203/02 *The British Horseracing Board Ltd and Others v William Hill Organization Ltd.*, para 87.

¹²⁶ CHIC Deliverable No. 8.2 Prototype implementation of the CHIC repositories, Section 5, p.20 et seq.

¹²⁷ *Ibid.*

use¹²⁸. The deployment of technological resources, knowledge and experience can be considered as a substantial investment in presenting the contents of the database. Such investment may be deemed sufficient to render the CDR protection by sui generis database rights. The technological background of constructing the CDR, structuring the data and presenting such data to the users is extensively described in the Deliverable No. 8.2 Prototype implementation of the CHIC repositories¹²⁹, which documents the investment made and can, under circumstances, be used to prove the protection of CDR by sui generis rights.

The sui generis database rights pass to the maker of the database, meaning “*the person who takes the initiative and the risk of investing*”¹³⁰. The party, who took the initiative and the risk in investing into the CDR is UBERN, who accordingly shall qualify to hold the sui generis database rights in the CDR. This will enable UBERN, both from technical and legal perspective, to be in the position to manage the use of the repository. The right holder may stipulate the terms of using the repository contents as a whole, grant the rights of use under contractual license, prevent and enforce the unauthorized extraction/reutilization of the repository contents, allocate the access rights to the project parties or external users, define the rights of use (access only, modification, download, etc.) and different regimes of uses depending on the data stored therein. The holder of sui generis database rights may leverage how the contents of its repository may be used, whether the data items may be extracted (downloaded) and in what scope, whether the data may be transferred to external parties or whether the data procession may only be done on the premises of CHIC.

For instance, UBERN has realized some of the rights of the right holder within the collaboration between CHIC and MyHealth Avatar (MHA)¹³¹. Under the Collaboration Agreement between CHIC and MyHealthAvatar, UBERN created for MHA a specific section in the CDR, allocated for storing the MHA data, and granted the MHA parties Access Rights to the CHIC data repository and application programming interfaces (access API) to access the repository, upload, download, store and retrieve data for MyHealthAvatar. Through the grant of these rights, legal and technological measures were put in place to limit the access of MHA parties to the dedicated repository section only without having access to the CHIC data in the CDR¹³².

However, the sui generis protection applies to the contents of the repository as a whole or in substantial part and may apply separately and irrespective of protectability of data items by other rights, such as copyrights. Article 7 (4) Database Directive makes this explicit, saying that the 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*”.

It follows that the holder of the repository may manage the use of the repository contents as a whole. However, the use of separate data items in the repository may remain governed by the terms, stipulated by the data providers and/or holders of rights in such items. For instance, the access rights to the datasets, handled as confidential, may require signing of a non-disclosure agreement (NDA) and the use of such data may be limited and be subject to technical protection measures, etc. The protectability of clinical data in CHIC under alternative regimes is considered in part 6.4 below.

¹²⁸ Ibid.

¹²⁹ Ibid.

¹³⁰ Recital 41 Database Directive.

¹³¹ MyHealthAvatar, full title “A Demonstration of 4D Digital Avatar Infrastructure for Access of Complete Patient Information, an FP7 funded project, Grant Agreement: 600929, available at: http://www.myhealthavatar.eu/?page_id=769.

¹³² CHIC-MHA Collaboration Agreement, Version 5, January 18, 2016.

6.3.3.2 Model/tool and in silico trial repositories

The Model/Tool Repository has been developed by ICCS and provides a software platform to store models (including hypomodels, hypermodels) and software tools, such as data transformation tools, linkers, tools used in the development and execution of models are stored¹³³. Via the graphical user interface, authorized end users (researchers, modellers, clinicians, etc.) may access the stored models/tools, store new ones, delete existing models/tools and update them¹³⁴.

The in silico trial repository, developed by ICCS, provides a technological platform to store the data related to in silico trials and in silico experiments. It is designed to host simulation scenarios and in silico predictions in a consistent manner. The repository is constructed in such a way that end users and other technological components in CHIC can access the repository in order to prepare new in silico trials and store the results of the conducted experiments.¹³⁵

Starting with qualification of the model/tool repository as a “database”, first, a database means “a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means.”¹³⁶ The contents of the model/tool repository is represented by computer models and software tools (which qualify as works protected by copyright), which are arranged in a way as to be accessible to the users. As follows from Figure 28 Entity-Relationship (ER) diagram of model/tool repository, provided in Deliverable No. 8.2 (Prototype implementation of the CHIC repositories), the models and tools are allocated according to certain parameters and properties so as to be identifiable and re-usable by the components in the CHIC hyper-modelling system. Against this background, the model/tool repository may be considered as a database in the meaning of Database Directive.

The in-silico trial repository consists of three main elements: the subjects, the experiments and the trials. The subject usually stands for a person, subject to the trial. The entity “experiment” consists of a triple: “initial state of the subject” – “in silico hyper model” – “final state of the subject”. All in silico experiments are organised into in silico trials. All in silico experiments that use the same hyper-model are in the same in silico trial¹³⁷. The elements, which inhabit the repository, may be the data, models, URL links, etc., which qualify as either works or data and can constitute the contents of the database, as defined by the Database Directive. This division of the repository contents as per entities means that the data is stored in the in silico repository in a structured way. Also, the entities: subject-experiment-trial are stored individually, so one and the same experiment can be viewed by different users. In the process the in silico trial repository collects data about subject, experiments and trials in a structured way, and makes the relevant datasets (per entity) accessible by electronic means. Thereby it fulfils the definition of a “database”.

In the next step, we need to ascertain whether either the collecting, verification or presentation of the contents of these repositories consumed the investment, which would justify the protection by sui generis database rights. In fact the requisite investment in collecting and verifying the models may be problematic, given that the models are provided by the modelling parties, and the validation of models occurs within the in silico trials. Thus there is no investment on the part of ICCS itself in collecting the data on subjects or experiments or in silico trials or verifying such data. What may instead justify sui generis protection for the ICCS repositories is the investment in presenting the repository contents.

¹³³ CHIC Deliverable D12.4 – Draft Plan for the Use and Dissemination of Foreground Section 3.1.2, p. 28.

¹³⁴ CHIC Deliverable No. 8.2 Prototype implementation of the CHIC repositories, Section 6, p.68 et seq.

¹³⁵ CHIC Deliverable No. 8.2, Section 6.1.

¹³⁶ Article 1 Database Directive.

¹³⁷ CHIC Deliverable D 8.1– Design of the CHIC repositories, Section 4.4, p.42.

The model/tool and in silico trial repositories are built on the same technological platforms. The deployment of the model/tool repository, and also in silico trial repository, is extensively described in Deliverable D8.2 (Prototype implementation of the CHIC repositories). As described there, the model/tool and in silico trial repository are constructed in compliance with the Django MVT pattern, chosen as a web application framework, and make use of Apache http server as an application server and rely on MySQL database server for storing the structured data. It should be noted, that the respective models/tools and in silico trials are stored in the model/tool and in silico trial databases. These databases rely on MySQL database architecture. However, for the models/tools or in silico experiments to be accessible to the end users (researchers, clinicians, administrator), the requests from the users to the repository need to pass through data transformation process. First, the requests for data from the repositories are accepted by the web server NginX. Then, the requests are divided into dynamic driven pages and static pages, the requests for dynamic pages are forwarded to the Apache server and handled by the Django repository framework¹³⁸.

This complex technological framework of deploying the repositories shows that the process of structuring and making the contents of the model/tool and in silico trial repositories accessible and re-usable to the users requires considerable computational power, technological resources, software and database architectures, knowledge and skill in order to combine these resources into a functional database operating from the web. The investment of such technological and human resources into constructing the repositories may be considered as substantial enough to justify sui generis protection.

ICCS is the party, who took the risk and initiative to invest into creating the model/tool and in silico trial repositories and, accordingly, will qualify as the database maker. Thus, ICCS, as the holder of database rights, shall have not only the technical possibility to administer the use and accessibility of the repositories, but also obtains the legally protected rights to prevent extraction and re-use of the contents from these repositories, if such acts affect the repository contents as a whole or in substantial part or interfere with the legitimate interests of the right holder¹³⁹. As maker of the databases, ICCS may transfer, assign and grant the rights on use of the repositories under contractual license, define the access and use rights to the CHIC parties and external users, apply technological measures to enforce the permissions in technical way and manage the use of the repositories at its discretion, but in the interest of the project and subject to the provisions of CHIC EC-GA and CA.

6.3.4 Sui generis database rights and rights in software and data

It must be noted, that protection of the repositories by sui generis database rights does not interfere and does not substitute the protection of software services and tools, used in construction of the repositories by software copyright. As expressly stated in the Database Directive¹⁴⁰, the term “database” shall not include “*computer programs used in the making or operation of a database*”. Computer programs, used in constructing the repository architectures (both as any other computer programs either used for operation of databases or not) are protected by software copyright¹⁴¹. For instance, MySQL, “*the world's most popular open source database*”¹⁴², which is used to store the structured data in the model/tool and in silico trial repositories is protected by software copyright and is licensed under open

¹³⁸ CHIC Deliverable No. 8.2, p. 71.

¹³⁹ Article 7 (1), (5) Database Directive.

¹⁴⁰ Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, OJ Nr. L 77/20, 27. 3. 96.

¹⁴¹ Directive 91/250/EEC of 14 May 1991 on the legal protection of computer programs, OJ No L 122, 17. 5. 1991, p.42. Directive as last amended by Directive 93/98/EEC (OJ No L 290, 24. 11. 1993, p. 9.)

¹⁴² <https://www.mysql.com/products/>

source software license GPL v2¹⁴³. Therefore, the software libraries, architectures, frameworks, which are used in constructing or operating the repositories are normally protected in their own right. The copyrights in programs or codes, which have been developed by the project parties, pass to these parties. The copyright in the third party libraries, architectures, frameworks, which were used in CHIC under “open source” schema, remain by the third parties.

Also, the sui generis protection does not preclude the elements, which are stored in the repositories, from being protected in their own right. Article 7 (4) Database Directive provides that sui generis database rights shall “*apply irrespective of the eligibility of that database for protection by copyright or by other rights. Moreover, it shall apply irrespective of eligibility of the contents of that database for protection by copyright or by other rights. Protection of databases under the right provided for in paragraph 1 shall be without prejudice to rights existing in respect of their contents.*” Thus, models stored in the model/tool repository constitute computer programs and are protected by software copyrights, results of in silico trials may be protected as R&D by the regime of undisclosed information, datasets in the CDR may also be protected in their own right. The potential regimes for protecting the clinical data by IPR and contractual schemes we consider next.

6.4 Protectability of clinical data by IP rights

The options of protecting the clinical data and data curation in CHIC by IP rights were explored in Deliverable D4.4 Whitepaper (Recommendations for an amended European legal framework on patients' and researchers' rights and duties in E-health related research). This included the protection offered by sui generis database rights, as applicable to the CDR, copyrights and related rights, and the application of legal regime of undisclosed information to protect the clinical data in CHIC.¹⁴⁴ The outcomes of that legal analysis may be summed up as follows:

- (a) The sui generis protection, applicable to the CDR, may be an option to protect the clinical data, stored in the CDR, from re-use or retrieval in whole or in substantial portion of the repository contents. However, the sui generis database rights protect the repository contents as a whole, and do not protect isolated datasets per se.
- (b) Protection by copyright requires expression of original intellectual creation in a work, which is normally not present in clinical data, coming from clinical trials or laboratory examinations, or clinical images, generated automatically. These are only reports, written by a physician or the patient, which expose the necessary level of original creativity, that may be protected by copyrights. In contrast, the clinical data in CHIC is normally expressed in numerical parameters, because only such data may be used for running the oncosimulations in CHIC. The images, processed in CHIC, are usually DICOM files, which are produced by automatic means and do not reflect personality of the author, which is a requisite for protection by image rights¹⁴⁵.
- (c) The number of related rights is as of now limited to phonograms, fixations of broadcasts, performances and films, fixations of audio-visual works and do not cover the types of data, produced in CHIC.
- (d) Protecting the clinical data by the legal regime of undisclosed information is possible, would require, however, adoption of additional organisational and technological measures.

¹⁴³ <http://dev.mysql.com/downloads/mysql/>

¹⁴⁴ Deliverable D4.4, section 6.2.

¹⁴⁵ CJEU, Judgment of 7 March 2013, Case C 145/10 REC, Eva-Maria Painer v. Standard VerlagsGmbH, Axel Springer AG, Süddeutsche Zeitung GmbH, Spiegel-Verlag Rudolf Augstein GmbH & Co. KG, Verlag M. DuMont Schauberg Expedition der Kölnischen Zeitung GmbH & Co. KG.

The possibility of protecting the clinical data in CHIC under the legal regime of undisclosed information and under contractual schemes is considered in more detail below. Naturally, where personal data (e.g. of patients) is involved, the feasibility of data IPR exploitation will in any case be curtailed by privacy and data protection rights; the analysis that follows is without prejudice to the application of these data subject rights.

6.4.1 Undisclosed information

In order to meet the criteria of “undisclosed information” and be protected as such, the data must be identifiable, it must be secret and have economic value¹⁴⁶. First, to satisfy the criterion of secrecy, the information, sought to be protected, must be accessible to a limited number of persons only. The use of such information must be subject to confidentiality measures. The application of confidentiality measures means that the data must be stamped as “Confidential” and the sharing of such data must be made upon non-disclosure obligation and observation of the confidentiality measures. Disclosure of such datasets without due confidentiality measures might compromise the regime of secrecy so that protection would be forfeited. As regards the requirement of economic value, this will be considered to be present if through publication of such data, the research investment and competitive standing of the entity doing the work would be undermined¹⁴⁷.

In principle, the clinical data in CHIC can be brought under the coverage of “undisclosed information”. However, considering the volumes of data, expression of such data in numerical parameters according to clinical protocols, possible availability of such data in clinical trial databases, it may be questioned whether seeking such protection for the whole amount of data is needed.

At the same time, the legal regime of undisclosed information would make a plausible and justifiable option for protecting some specified datasets. For this, the data items, selected for protection, must be identified, the access and use of such data shall be limited to a defined number of people only, the management of such data shall be subject to confidentiality measures. In the case of CHIC, the regime of secrecy may be provided to the data via marking it as “confidential”. Considering from the technical side, the confidentiality mark would then need to be placed and borne by the data throughout the whole research process so that the data marked as “confidential” by the input comes out marked “confidential” by the output. This would present an additional workload, but is implementable. Also, disclosure of such data items to the CHIC parties subject to the non-disclosure obligation would not present a significant obstacle, because the project parties are bound by the contractual relations within the project. By being marked as “confidential”, the data would automatically fall under the regime non-disclosure of information, as provided by Section 10 CHIC CA. Accordingly, all CHIC parties, who get access to such data, are bound by the non-disclosure obligation and must treat this data accordingly. The factual use of data within the project may also be managed by technical measures, such as granting or denying the access rights, rights of use and extraction, limiting the data procession to the framework of technical infrastructure of CHIC only.

Whereas the application of such contractual and technical confidentiality measures to the clinical data in CHIC may be feasible, in how far such technical and confidentiality measures may be implemented and in how far the control over the use of such data may be exercised by release of the data outside the domain of CHIC is questionable. And even the adoption of technical and legal measures to preserve the confidentiality of data and release of such data upon signature of non-disclosure agreement would not provide adequate protection, once the data leaves the contractual, technological and the security framework of CHIC.

¹⁴⁶ Article 39 TRIPS Agreement.

¹⁴⁷ K. Lodigkeit, Intellectual Property Rights in Computer Programs in the USA and Germany, Peter Lang GmbH, 2006, pp. 98-101.

6.4.2 Contractual approaches

Apart from the applicability of the confidentiality measures to certain data, also release of some data sets and the use of such data both within and beyond the domain of CHIC may be constrained by contractual terms. Even where the data is not supposed to be treated as “confidential”, the data providers may nevertheless stipulate certain terms, under which such data may and must be used. For instance, the data provider may wish to release the data for non-commercial research only and forbid any commercial use, or constrain the use of the data within the VPH modelling community only, or identify his name as data provider, etc. The application of contractual regimes and constraining the use of the data within contractual terms is possible and technically implementable.

However, weak points in applying contractual regimes in relation to data are: loss of control over the use of data after it leaves the premises of data provider/CHIC, the necessity to conclude the agreements in writing for the contractual claims to be enforceable in courts, an indefinite number of potential users, and non-reliability of contractual parties. On the basis of these observations, the application of contractual approaches in relation to the clinical data may be deemed practicable within the project and within a limited number of reliable parties, but does not make a reasonable option for releasing the data into use to the broader community.

6.4.3 Data ownership issues

In principle, the ownership issues within the CHIC project are governed by the rules of the CHIC CA. Article 8.1 CHIC CA provides: *“Foreground shall be owned by the Party who carried out the work generating the Foreground, or on whose behalf such work was carried out”*. This approach may well be applicable to defining ownership in the software components or models. Following this rule, a party, who created software, implementing one or another model (mechanistic model, vasculature, bio-mechanic model, etc.) shall own the copyrights in such software and the corresponding computer model, respectively. However, this rule may not be as directly applicable to the ownership in data outputs, produced by the models.

Starting with the definition of foreground first: in the meaning of Article 8.1 CHIC CA and the legal framework of CHIC in general, the foreground means *“the results, including information, whether or not they can be protected, which are generated under the project. Such results include rights related to copyright; design rights; patent rights; plant variety rights; or similar forms of protection”*¹⁴⁸. Hence, the data outputs, produced by the models, constitute information generated under the CHIC project and fall under the definition of Foreground, as defined by the legal framework of CHIC. However, although generated under the project and using the research infrastructure of CHIC, such data outputs cannot be directly covered by rules of defining ownership, as provided for by the CHIC CA.

First, the data outputs from computer models are very likely not to be protected by IP rights, except specific cases where confidentiality measures have been applied. The data outputs constitute the results of simulations, run by the models. Such data outputs are produced by means of computer software and lack any intellectual input to be covered by copyrights. Unless confidentiality measures and “Confidential” marks are applied, such data outputs would remain outside the scope of the legal regime of undisclosed information, provided by Section 10 CA, and the scope of rights in such information as well.

Second, the ownership over the data outputs, generated by the models, is not directly covered by the rule on the ownership in Foreground, as stipulated by the CHIC CA. As noted above, Article 8.1 CA attributes ownership in foreground to the party, who carried out the work, generating such foreground. Following this rule, there are two potential candidates, who may claim ownership in data outputs, namely: the modelling parties and the clinical parties. Both the modelling and the clinical parties carried out the work, leading to the

¹⁴⁸ Article II.1, Annex II, EC-GA.

generation of model outputs, but neither modelling not clinical parties produced such data outputs themselves.

Following the rule of Article 8.1 CA, the modelling parties may argue that they carried out the work generating the models, which perform the simulations and calculate the data, and therefore they shall own the data outputs, produced in the result of execution of their models. The data outputs constitute the results of data procession, realized by the models and the modelling parties, who own the models, may claim to own the rights in the data outputs obtained in the result of data procession done by their models. Article 950 (1) of German Civil Code¹⁴⁹ may also be brought to consideration here. It provides: *“A person who, by processing or transformation of one or more substances, creates a new movable thing acquires the ownership of the new thing, except where the value of the processing or the transformation is substantially less than the value of the substance. Processing also includes writing, drawing, painting, printing, engraving or a similar processing of the surface.”* In how far this provision may be applicable to data, which constitute intangible assets, in contrast to movable things, is rather questionable and subject to interpretation of German civil law¹⁵⁰. Again, whether the term “movable thing” and ownership over a movable thing can be applied in relation to data is subject to the legal debate¹⁵¹.

In their turn, though, the clinical parties might argue that they carried out the work generating the raw clinical data, which was used by the models as inputs to produce the outputs. They may support this by analogy to the fact that when making calculations by MS Excel does not allocate any rights in the data outputs to Microsoft. In contrast, he, who computes the data using MS Excel, normally does so under a MS Excel software license and pays Microsoft license fees for having the right to use MS Excel. In any event, as this discussion shows, the issues surrounding data IPR exploitability (even aside from data protection aspects adverted to earlier) remain highly complex and problematic; here the more recognised IPR modalities, discussed earlier offer a preferable means for parties to protect their project investment.

¹⁴⁹ Civil Code in the version promulgated on 2 January 2002 (Federal Law Gazette [*Bundesgesetzblatt*] I page 42, 2909; 2003 I page 738), last amended by Article 4 para. 5 of the Act of 1 October 2013 (Federal Law Gazette I page 3719).

¹⁵⁰ Article 90 of the German Civil Code (Bürgerliches Gesetzbuch).

¹⁵¹ Barbara J Evans, ‘Much ado about data ownership’, (2011) 25 Harv. J. L. & Tech. 78, p. 69-130.

7. Conclusion

As appears from the foregoing chapters, a number of factors have needed to be taken account of in evolving the CHIC second iteration data protection and copyright framework. Thus the requirements for processing sensitive health data under the European legal regime have required application in arriving at a workable solution, where cancer modellers, and clinical oncologists treating patients, are able to process patient data for validating the cancer models developed during the project. This framework in continuity with the first iteration framework seeks to ensure the highest level of data protection while maintaining this flexibility. In this regard, the safety net devised in the first iteration, using strong technical, organisational and legal measures to ensure data is processed within a context of anonymity has been continued, in particular to cover the feed-back by validating clinicians of model predictions and patient outcomes to the CHIC modelers. These measures aim to ensure an appropriate level of security of the data processing, taking into account the state of the art and the costs of their implementation relative to the risks inherent in the processing and the nature of the data to be protected.

The contractual obligations in the appended agreements to this deliverable (Appendices 3-5) highlight the commitments of the relevant parties to look after the data appropriately, and use it only as strictly required for achieving the validation and adaptation of the models. However, as a precondition for the clinicians to use of prospective patient data when validating models, and supplemental to the need for strict data protection requirements, it will be essential to take due account of the patient's fundamental interest in dignity and autonomy. To this end an informative and clearly-phrased model patient information sheet and consent, to inform the patient of the implications of participating in the model validation process has been drafted, and is presented in Appendix 2.

Most legal issues with respect to the management of IPR in CHIC have been settled in the IPR Memorandum of Understanding, signed and entered into force on 30 June 2015. The legal analysis revealed that the CHIC repositories, i.e. the clinical data repository (UBERN), the model/tool repository and in silico trial repositories are likely to be protected by sui generis database rights, allowing the makers of these repositories to prevent unauthorized extraction and reuse of the contents of their repositories in whole or in substantial part. However, the protection, provided by sui generis database rights covers the repository contents as a whole and does not interfere and may co-exist with the protection of the repository architecture by software copyrights and protectability of the individual elements of the contents by other legal rights, such as copyrights.

Among the options for protection (i.e., copyrights, related rights, database rights, undisclosed information) the sui generis database rights and the legal regime of undisclosed information constitute practicable solutions. Thus, the clinical data collected in the clinical data repository is to be covered by the sui generis database rights, applicable to the repository. However, such protection covers the repository contents as a whole, and does not extend to individual datasets. Some specific datasets, which are considered to be of a potentially higher value or sensitivity (such as genetic data), may, in supplement to the data protection and security mechanisms, be handled as "confidential". Bearing such marking, such datasets will fall under the legal regime of undisclosed information and will be treated and protected as such (Section 10 CA). In addition future exploitation risks stemming from the potential incompatibility of software licenses applicable to the diverse CHIC models and tools developed in the project have been addressed through an analysis, with the results presented in the CHIC software licensing report in Appendix 6.

8 References

- Article 29 Working Party, Opinion 4/2007 on the Concept of Personal Data, adopted on 20.06.2007.
- Article 29 Working Party, Opinion 15/2011 on the definition of consent; WP187
- Article 29 Working Party, Opinion 05/2014 on Anonymization Techniques, adopted on 10.04.2014.
- Ciriani, V. De Capitani di Vimercati, S. Foresti, S and Samarati, P. k-Anonymity
- Davison, I, The Legal Protection of Data Bases
- Declaration of Helsinki, World Medical Association, 1964 (latest revision, Brazil, 2013)
- Derclaye, E., *The Legal Protection of Databases*, (2008).
- El Emam, K., and Fida Kamal Dankar, F., Protecting Privacy Using k-Anonymity, 2008.
- Evans, B.J., ‘Much ado about data ownership’, (2011) 25 Harv. J. L. & Tech. 78
- Forgó, N, et.al., Ethical and Legal Requirements for Transnational Genetic Research, 2010
- Garfinkel, S., “De-Identification of Personally Identifiable Information”, DRAFT NISTIR 8053.
- HITRUST, HITRUST De-Identification Framework
- ICO, Anonymization: Managing Data Protection Risk Code of Practice, November 2012
- Lee Hansen, S, Mukherjee, S., A Polynomial Algorithm for Optimal Microaggregation, 2003
- LeFevre, K, DeWitt, D.J, Ramakrishnan, R., Incognito: Efficient Full Domain K-Anonymity, 2005.
- Lodigkeit, K., Intellectual Property Rights in Computer Programs in the USA and Germany, Peter Lang GmbH, 2006
- Ninghui Li, Tiancheng Li, Suresh Venkatasubramanian. t-Closeness: Privacy Beyond k-Anonymity and l-Diversity. ICDE, IEEE 2007, 106.
- Novartis Deutschland GmbH, Novartis Global Data Anonymization Standards
- Phuse De-Identification Working Group, De-Identification Standard for CDISC SDTM 3.2, version 1.01, 20.05.2015
- Project Data Sphere, LLC, A De-Identification Strategy Used for Sharing One Data Provider’s Oncology Trials Data through the Project Data Sphere® Repository
- Rowland, D., Kohl, U., and Charlesworth, A., *Information Technology Law*, 2012.
- St. Laurent, A.M., “Understanding Open Source and Free Software Licensing”, 2004.

Tiancheng Li, Ninghui Li, Jian Zhang, Molloy, I., Slicing: A new approach to privacy preserving data publishing. 2009

TransCelerate Biopharma Inc., Data de-identification and Anonymization of Individual Patient Data in Clinical Studies – A Model Approach (2015)

Xiaoxun Sun, Min Li, Hua Wang, Plank, A., An efficient hash-based algorithm for minimal k-anonymity, 2008.

Appendix 1 – Abbreviations and acronyms

<i>CA</i>	Consortium Agreement
<i>CHIC</i>	Computational Horizon in Cancer
<i>CATS</i>	Custodix Anonymisation Tool Services
<i>CDP</i>	Center for Data Protection
<i>DoW</i>	Description of Work
<i>EC</i>	European Commission
<i>ECJ</i>	European Court of Justice
<i>GDPR</i>	General Data Protection Regulation
<i>GPL</i>	General Public License
<i>GBM</i>	Glioblastoma Multiforme
<i>HSCIC</i>	Health and Social Care Information Centre
<i>IAM</i>	Identity and Access Management
<i>IdP</i>	Identity Provider
<i>IPR</i>	Intellectual Property Rights
<i>ICT</i>	Information and Communication Technology
<i>LGPL</i>	Library General Public License
<i>NSCLC</i>	Non Small Cell Lung Cancer
<i>PDP</i>	Policy Decision Point
<i>PAP</i>	Policy Administration Point
<i>PEP</i>	Policy Enforcement Point
<i>PIMS</i>	Patient Identity Management System
<i>PIP</i>	Policy Information Point
<i>SAML</i>	Security Assertion Markup Language
<i>STS</i>	Secure Token Service
<i>TTP</i>	Trusted Third Party
<i>TSP</i>	Health and Social Care Information Centre
<i>TRIPS</i>	Agreement on Trade-Related Aspects of Intellectual Property Rights
<i>VHP</i>	Virtual Physiological Human
<i>WIPO</i>	World Intellectual Property Organisation

Appendix 2 – Patient Consent to Data Use for Model Validation

Patient information sheet (Version.01; September. 2016)

Explanation of the CHIC (“Computational Horizons in Cancer”) validation study

Invitation:

You are being invited to take part in an observational clinical validation study involving patients and their health data. The purpose of the study is for clinicians to carry out early assessment of a new clinical decision support tool that is designed to provide them with information to further optimize the treatment they provide.

This document describes the study in order to help you to make your decision. Please read the information provided carefully and discuss it with others if you wish. Feel free to ask your medical doctor and other members of your healthcare team if there is anything that is unclear or if you would like more information.

Please take time to decide whether or not you wish to take part. You must not feel obliged to participate in this study. If you do decide to participate, you can withdraw your consent at any time without any disadvantages. Also, if you decide not to volunteer for the study, it will not affect your treatment in any way.

Thank you for reading this.

[1] Purpose of the CHIC project and validation study:

The Computational Horizons In Cancer (CHIC) project is a EU-financed FP7 project that aims to contribute to the advance of so-called personalised medicine in cancer care. More specifically, it will establish an infrastructure for the development of predictive computational cancer models that allow cancer treatment to be tailored to the characteristics of each patient. The models seek to synthesise medical and scientific knowledge as to how tumours develop or decrease under many different conditions (including diverse therapies and specific genomic or other patient characteristics). The idea is for clinicians then to use them, populated with the data of a particular patient, to generate a (probabilistic) prediction of the form of therapy likely to produce the best medical outcome for that patient.

The project, which began in April 2013, has now reached the point where, for certain cancers, models are available for initial clinical testing by clinicians. At this stage, the clinicians will be seeking to assess the broad usability of the models in clinical practice (in terms of their design for inputting data and presenting back results), as well as to gain an early sense of whether the treatment predictions and recommendations made by the models offer a useful supplement to the clinician's own judgment based on his or her medical expertise and previous experience in treating that form of cancer.

It is important to emphasize that during the study the clinician will not rely upon any predictions made by the models. This is because at present the models are still in an immature state, where their accuracy in clinical practice has yet to be tested or proven. Rather, the preliminary assessment that clinicians are now carrying out will be the first step on a path that it is hoped will eventually lead to the wider approval and adoption of such computational models in the provision of personalized cancer care.

If you decide to take part in the CHIC project study, you will be asked to agree to your clinician transferring your personal health data (including data collected for any other clinical trial) to the CHIC project platform, from where the clinician will access and use the data to

run one or more of the computerized cancer models so as to generate an individual treatment prediction for your case. For this purpose your data will be stored in the CHIC platform in pseudonymised (key-coded) form, so that your clinician – who has the key - can keep track of the data and the model's prediction for you. As stated, your clinician will not make use of this information when determining your therapy; however, he or she will retain it for comparison against the outcome achieved following the actual chosen therapy. Later, if you agree, the clinician may also supply both sets of information (i.e. the model prediction and the actual outcome) in securely de-identified form to the IT scientists who developed the relevant models. This could help them to further adapt and refine the models.

[2] Why have I been invited to make my data available for the CHIC validation study?

You have been invited to participate in the CHIC validation study for two main reasons. First, you have been diagnosed with one of the forms of cancer, namely [.....] for which the CHIC project has developed one or more computational predictive models that are ready for preliminary validation in clinical practice. Secondly, your treating clinician [name of clinician] has agreed to act as one of the clinical expert model validators. In this regard, you are being invited together with other patients at [name of hospital] under the care of [name of clinician] to allow him/her to transfer your personal health data to the CHIC project platform to use it for validating the relevant models.

Your health data will be stored securely within the CHIC platform for potential use at two points in time and for two distinct purposes. First, your clinician will use the data to run a computational model, yielding an individual prediction of the potential outcomes for you from receiving different therapies; this information will only be seen by your clinician and will help him form an impression of the model's usefulness for clinical decision support. Second, your clinician may (subject to your separate agreement) later share the data concerning the model predictions, plus data recording your actual treatment progress, in de-identified form - and still within the CHIC platform - with the IT scientists who developed the relevant models, for use in refining these further.

[3] Do I have to take part in the CHIC validation study?

Your decision to make your data available to the CHIC platform for the validation study is entirely voluntary. If you decide to take part you will be asked to sign the consent form below. By signing the consent form, you will confirm you were properly informed about this study and that all your questions have been answered. A copy of the patient information sheet and of the consent form will be given to you to keep.

In this case, you will remain free to withdraw your consent at any time, and have your health data permanently deleted from the CHIC platform without giving any reason. A decision to withdraw, or not to take part in the CHIC validation study in the first place, will in no way affect your medical care or the relationship with your clinician or other medical staff.

[4] How is my data protected?

Your health data will initially be transferred and stored in a secure database within the CHIC platform, to which only your clinician has access. The data will be stored in pseudonymised form (where your name and other personal identifiers such as address and date of birth) are replaced by a key-code; this allows your clinician alone – who has the key – to easily relate the data to you. After the clinician uses the data to run a computational model, the outcome data, i.e. the model's prediction of the possible treatment outcomes, will also be stored in this way. The CHIC platform as a whole is protected by state of the art security against unauthorized access, including strict user authentication, firewalls, and data encryption technology. Later, if you agree to making your data available to the IT scientists to help refine the models, it will be subject to further secure de-identification (double-pseudonymisation, plus removal of sensitive data values such as real dates; the IT scientists will also be contractually bound not to disclose the data further and to keep it secure at all times.

Your data will only be stored in the CHIC platform for as long as it is required in order to validate and/or refine the CHIC cancer models. After that it will be irreversibly deleted. If, at an earlier time you choose to withdraw your consent to further storage, your data will then be irreversibly deleted and you will receive a confirmation from your clinician that this has occurred.

[5] Will I be informed about the results of the validation study?

As explained under [1] above, your clinician will as part of the CHIC validation study receive an individuated computational model prediction as to your likely response to different potential cancer therapies. However, as has also been noted, the information at issue is probabilistic and especially at this early stage its relative degree of accuracy remains unknown. For this reason the clinician will not place any weight on the model when selecting your treatment, but instead rely upon his or her independent clinical judgment. Similarly, the decision whether, or how far, to divulge or discuss the model's prediction with you, the patient, will principally be for the clinician, in line with his or her individual clinical discretion.

The clinician would naturally take strong account of your wishes, and in the consent below you are invited to signal what your attitude towards receiving the information might be. However, the clinician would also not disclose information, unless personally satisfied that the model is reasonably accurate and the information otherwise clinically relevant. As regards general information about the progress and results of the CHIC validation study, you are welcome to ask your clinician, who will be pleased to provide details of where to find this.

[6] Risks and benefits of this project

You are very unlikely to receive a direct treatment benefit from agreeing to take part in the validation study at this stage. There is a possibility that a model may, in a particular case, yield information of special clinical interest, where there are sound independent reasons for thinking the information reliable. In that case the treating clinician, in discussion with the patient, would decide how best to make use of the information. However, such cases are expected to be rare. Instead, the main benefit may be the knowledge that, in taking part in the CHIC validation study with its aim of improving cancer clinical decision support, you are helping to advance medical knowledge, and the development of new and better treatment options for patients in the future.

As explained under [4] above, the privacy of your data is guaranteed by its secure encrypted storage within the CHIC platform, subject to strict access control so that in the first instance only your clinician may see and use the data; and secondly (if you agree) the IT scientists, but only to the data in securely de-identified form and subject to a strict legal duty of care. There are no other risks anticipated from your involvement in the study.

[7] Costs

There will not be any additional costs for you if you decide to participate in the CHIC validation study.

[The Consent Form for signature appears on the following page.]

Consent Form for taking part in the CHIC validation study (Version .01; September 2016)

I, the undersigned, born on the....., in..... and resident at /(address),

Declare that I have read and understand the Patient information sheet - which form a part of this document (version .01; September 2016))

And that I agree for my treating clinician..... at the hospital to make use of my health data for the purpose of validating CHIC cancer models in the CHIC ("Computational Horizons in Cancer") validation study.

I further agree to the subsequent use of my data in securely de-identified form by the IT scientists who developed the relevant models for the purpose of further refining and improving the said models: [Patient to initial separately if applicable]

I should be interested to learn from and discuss with my clinician the nature of the prediction yielded by the computational model on the basis of my health data: Yes, If my clinician believes the information relevant for my treatment..... / No..... [Tick as applicable].

I understand that two original copies of this agreement will be produced and retained by myself and the [Name of hospital].

Name of the patient:

Name of patient legal representative (if applicable):

Name of treating clinician:

Name of hospital representative:

Signature of the patient (or if applicable, the patient's legal representative):

.....

Date (please date your own signature):

Appendix 3 – CHIC Clinical User Agreement

(Version 2.0, September 2016)

BETWEEN

(“CHIC clinical user”)

(address and country of establishment)

AND

(“CDP/CHIC platform administrator”)

(address)

Individually referred to as a “Party” or collectively referred to as the “Parties”.

Preamble

The Computational Horizons In Cancer (CHIC) project is a EU-financed FP7 project that aims to create an infrastructure for the development of a number of integrative multiscale cancer models and hypermodel oncosimulators. These will be clinically adapted and partly validated, a process which will involve sharing of clinical and genomic data of patients within the project. At the same time, each of the partners recognises as a priority the imperative need to respect the fundamental interests and rights of patients, including the need to preserve the security and privacy of personal data involved in the project.

Therefore the infrastructure of CHIC is embedded in the CHIC Data Protection Framework, which guarantees compliance with current European data protection legislation, primarily by de facto anonymising the patient data. Due to the diverse participation of researchers in the project, it is of high importance to process patient data in compliance with all applicable laws and regulations, including without limitation, privacy and medical secrecy laws applicable to the activities of the parties. During the CHIC project, the CHIC platform infrastructure and security is administered by the CDP; after the end of the project, the intention is for this role to be taken over by the CHIC platform administrator.

To fulfil the objectives of the project, the patient data used by the modeller partners to build and adapt the models is subject to de-identification, using secure state of the art pseudonymisation/de-identification tool (eg, CATS), and stored in secure, access-controlled data repositories within the CHIC platform infrastructure. The CDP/CHIC platform administrator will transfer the original (data provider) code to an independent trusted third party, and the latter alone will retain the pseudonymisation key (cross table) needed to link the double-encrypted CHIC data set to the initial de-identified data sets provided by the clinical partners. However, in the current stage of the project, it is also important for the CHIC clinical users (clinical partners and individual clinicians in their employ) to begin validating the models in small numbers of prospective individual patients. For this purpose the

individual clinicians will be permitted to transfer to and access within the CHIC infrastructure the data of their own patients in single-pseudonymised form so that they are able to compare predictions generated by the models in respect of those patients against the actual treatment outcomes achieved. The CHIC clinical users will ensure that before doing so they obtain fully informed consent from the patients concerned and relevant ethics and other regulatory body approvals data providers; in addition they will adhere fully to the requirements of data protection, processing the patient data within the CHIC infrastructure subject to necessary technical and organisational safeguards to protect the data.

This agreement is needed to state the obligations and conditions under which a CHIC clinical user will transfer data to and process data within the CHIC infrastructure.

Clause 1: Definitions

For the purposes of this agreement, the terms used in these clauses shall have the same meaning as attributed to them in the Explanatory Glossary in Annex A to this agreement.

Clause 2: Scope and responsibility

1. This agreement sets out the terms and conditions for the transfer of patient data to and processing of the same within the CHIC infrastructure by the CHIC clinical user for the purposes of the validating the CHIC models.
2. The CHIC clinical user is responsible as data controller for the management of patient data within its organisation/ hospital database, while the CDP/CHIC administrator is responsible for the security of data following their transfer to the CHIC infrastructure.

Clause 3: Obligations of the CHIC clinical user

The **CHIC clinical user** warrants and undertakes:

1. to transfer to the CHIC platform infrastructure only data that have been collected and processed in accordance with the laws applicable to the clinical user;
2. that it shall obtain the consent of each patient to the use of the data in accordance with applicable ethical and legal norms together with the approval of its responsible ethics committee for the model validation process conducted using the data.
3. that it shall consult and provide notification as required to medical device regulatory authorities and national data protection authorities before transferring the data to the CHIC platform;
4. that it shall fully indemnify and hold harmless the CDP/CHIC platform administrator in case of any breach resulting from non-compliance with the relevant laws, including national data protection law applicable to the data transfer. Neither the CHIC Consortium nor the CDP will be liable in case of any issues arising from any deficiency in the consent of the patients, or non-notification of relevant authorities or local ethics bodies;
5. that initially, when transferring patient data to the CHIC platform for its own use in validating the CHIC models, it shall perform a thorough pseudonymisation process on the data so that all direct identifiers (patient name, address, date of birth, etc) are replaced by a code to which only the CHIC clinical user itself has the key;
6. that subsequently, when transferring patient data to the CHIC platform for the use of the CHIC modellers to adapt and refine the models, and provided always that the patient has consented to such further transfer, it will submit the data to a secure de-identification process using a state of the art de-identification tool in converting the personal data into de-identified data, so that the data as then accessed and used by the CHIC modellers may no longer be regarded as personal data. The assessment whether data has been properly de-identified remains with the CDP/CHIC platform administrator.
7. to support the CDP/CHIC platform administrator by if necessary performing further de-identification measures as reasonably directed by it, and by providing all necessary information and documents that may be needed in case of any request by supervisory authorities.

Clause 4: Obligations of the CDP/CHIC platform administrator

The **CDP/CHIC platform administrator** warrants and undertakes:

1. to have in place and implement appropriate technical and organisational security measures to protect the patient data against misuse and loss (including without limitation the measures stated in Annex B to this agreement) in accordance with the requirements of relevant provisions of European data protection law, in particular Article 17 of the Data Protection Directive 95/46/EC or Article 32 of the General Data Protection Regulation 2016/679 once this shall take effect.
2. to maintain the transferred data in a de-identified state, protected by secure encryption and other security measures as further detailed in Annex B to this agreement, such that it is not reasonably possible either for the CDP/CHIC platform administrator or any CHIC modeller user to relate the data to the original patient data subject;
3. to conclude binding contracts with the CHIC modeller end users (in the form of the CHIC end user agreement) in order to secure that any authorised CHIC end user which has access to the transferred data respects and maintains the confidentiality and security of the data;
4. to irreversibly delete the patient data from the CHIC platform infrastructure once it is no longer required for the purposes of developing or validating the CHIC models, or upon being informed by the CHIC clinical user that the patient no longer wishes his or her data to be used for these purposes (whichever is earlier) and to confirm to the CHIC clinical user that the deletion has occurred.
5. to comply with data protection laws applicable to its operations as well as the conditions set forth in this agreement;
6. to support the CHIC clinical data user by providing all necessary information and documents that may be needed in case of any request by supervisory authorities.
7. to deposit a copy of this agreement with the supervisory authority if it so requests or if such deposit is required under the applicable regulation.

Clause 5 Further Obligations of CHIC Platform Administrator

So far as it requires external support for data security to protect data processed after the end of the project, the CHIC Platform Administrator shall enter the 'CHIC Data Security Agreement' with the CHIC data security provider. It shall further ensure that each of its employees who has contact with the patient data is made aware of, and will be bound by, the terms of this Agreement, and such an employee will complete Annex C to this Agreement.

Clause 6: Liability and indemnity

1. Each Party shall be liable to the other Party for damage it causes by any breach of these clauses. The Parties agree that if one Party is held liable for a violation of the clauses committed by the other Party, the latter will, to the extent to which it is liable, indemnify the first party for any cost, charge, damages, expenses or loss it has incurred. Indemnification is contingent upon:
 - the Parties promptly notifying each other of a claim; and
 - each Party is given the possibility to cooperate in the defence and settlement of the claim.
2. The Parties agree that each Party shall be liable to the patient for damage caused by any wilful or negligent violation by it of data protection legislation or any analogous provisions of national or international law. The enforcement of this clause shall be subject to the finding of wilfulness or negligence by the court under Clause 8 below.
3. The above provisions shall be without prejudice to the Parties' right to terminate the contract, to seek compensation for damages or to enforce any claims under this agreement.

Clause 7: Termination and obligations of the parties after the termination

1. This agreement where entered into by the CDP shall terminate, unless superseded or amended by new provisions extending it, at the latest by 31 March 2017; where entered into by the CHIC Platform Administrator as successor to the CDP it shall terminate (unless extended) by..... [Date].
2. In case of breach of clauses 3 or 4 by one of the Parties, the other party is entitled to give written notice requiring the party in breach to be repair the breach within 72 hours, after which time if the breach remains outstanding it may terminate this agreement.
3. Without prejudice to the foregoing provisions, each Party may terminate this agreement for good cause.
4. Each Party must inform the other Party in written form in case of termination of the agreement.

Clause 8: Governing law and Jurisdiction, miscellaneous

1. This agreement shall be governed by Belgian Law. The courts of Brussels/Belgium shall have exclusive jurisdiction. This shall also apply to disputes on the validity of this clause.
2. Changes and amendments to this agreement shall require written agreement signed by the parties and an explicit statement that they represent a change or amendment to these conditions. The same applies to the waiving of this formal requirement.
3. If any provision of this agreement shall be entirely or partly invalid or unenforceable, this shall not affect the validity and enforceability of any other provision. An invalid or unenforceable provision shall be regarded as replaced by such a valid and enforceable provision that as closely as possible reflects the privacy/security and/or economic purpose that the Parties hereto had purposed with the invalid or unenforceable provision.
4. Each person signing below and each Party on whose behalf such person executes this agreement warrants that he/she, as the case may be, has the authority and the legal capacity to enter into this contractual agreement and perform the obligation herein.
5. This agreement will enter into force on the effective date, i.e. the date of the last binding signature to this agreement.

MADE in two signed copies, each Party having received its own signed copy.

(Place, Date)

(Auth'd Rep. of the CDP/CHIC platform administrator)

(Place, Date)

(Auth'd Rep of the CHIC clinical user)

Annexes:

- A. Explanatory Glossary
- B. Technical and organisational measures
- C. Access authentication form

[The Annexes appear on the following pages.]

Annex A: Explanatory Glossary

(Version 2.0, September 2016)

Clinician

The clinician is the natural person who is in charge of the patient's treatment.

Confidentiality duty

Persons engaged in data processing within the CHIC project and the CHIC platform infrastructure shall not, without authorisation, collect or process personal data, nor publish or disclose such data to any third party. The CHIC clinicians who provide and use the personal data of their patients are under a professional obligation of confidentiality towards them. Other parties, including the CHIC platform administrator and modeller users, that process data within the CHIC platform infrastructure shall require any of their staff with contact to the data to give an undertaking to maintain confidentiality as set out in of Annex C of this agreement. This undertaking shall continue to be valid after termination of their activity.

Consent

Informed consent means any express indication of the patient data subject's wishes, expressing his/her agreement to data relating to him/her being processed, provided that he/she has sufficient information about the purposes of the processing, the data or categories of data concerned, the recipient of the data, and the name and address of the controller and of his/her legal representative if any. The consent must be freely given and specific, and may be withdrawn by the subject at any time. If the subject is incapable of a free decision or domestic laws do not permit the subject to act on his/her own behalf, consent is required of the person recognised as legally entitled to act in the interest of the data subject or of an authority or any person or body provided for by law (legal representative).

Data controller

The data controller/controller is, according to the Data Protection Directive 95/46/EC, the natural or legal person who alone, or jointly with others, determines the purposes and means of the processing of personal data. The data controller is liable for the legality of the processing and the fulfilment of the obligations towards the national data protection authority and the patients. The CHIC clinical users and CDP/CHIC platform administrator are joint data controllers with regard to the collection and processing of the data in the CHIC platform infrastructure.

Data subject

The data subject is the subject of personal data, meaning an identified or identifiable person the data refers to. 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/her physical, physiological, mental, economic, cultural or social identity. A patient whose data is transferred by the CHIC clinical user in pseudonymised form for the clinical user to execute a model and receive back a prediction for that patient will be a data subject when his/her personal data are processed.

Disclosing

Disclosure is a processing operation in which patient and subject data are provided by a controller to a third party. The data controller must only disclose data to third parties if permitted by law or by the data subject's consent. The framework of agreements covering the CHIC model building and validation ensures that data are only shared with model end users who have each signed a contract that forbids any disclosure of data received via the CHIC platform to any third party.

Legal representative of the patient ("legal representative"):

The legal representative(s) of the patient is/are the person(s) who has/have the power by law or legal decision to decide for a minor patient (or equivalent status such as mentally disabled patients).

Necessary processing

When deciding which data will be collected and further processed, the controller must limit these data to the extent necessary to achieve the purpose of processing. This means that personal data will only be processed when it is necessary for the purposes of the CHIC project and CHIC model validation.

Patient:

Patient means the person treated in a clinic or hospital. Certain data collected by the CHIC clinical users will upon the patient's consent and the obtaining of appropriate ethics body approval be transferred to the CHIC platform infrastructure.

Personal data

Personal data means any 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/her physical, physiological, mental, economic, cultural or social identity. Therefore a set of data collected under a certain number or sign "patient xxx", "tissue YYY" can be personal data, if the patient concerned can still be identified by other means than his/her name.

Purpose

The purposes for processing of personal data must be adequate, relevant and not excessive in relation to the purposes for which they are collected and/or further processed. The purposes must be specified, explicit and legitimate. Personal data must not be further processed in a way incompatible with those purposes. The purpose for the collection, transfer and use of the data within the CHIC platform infrastructure is to create and validate predictive cancer models that are designed to assist clinical oncologists in their clinical decision-making.

Secure Deidentification

To securely deidentify a data set means to follow a state of the art de-identification process, which precludes re-identification of the patient / subject or makes it disproportionately difficult in terms of amount of time, expense and labour to re-attribute the data. Data de-identified in this way may be regarded as anonymous and no longer "personal data" in the legal sense. However it will still always be necessary to maintain strict data security against potential attempts to access and re-link the data in new ways, thereby re-rendering it personal data. In the CHIC platform infrastructure, it is an aim to have as much securely de-identified data as possible and reasonable, and to implement ongoing technical and organisation measures to secure the data (including the present framework of contractual data sharing agreements).

For the purposes of the CHIC model validation, the clinical users will carry out initial pseudonymisation on-site before sending the patient data to the CHIC platform infrastructure. When these clinical users themselves access and process the data, this will remain personal data in their hands (see 'personal data' above). This is justified according to the need for them to know who the patient (who remains in their care) is in order to assess the way the models perform. However, later if the clinicians transfer the data for use by the CHIC modeller users it will undergo further secure de-identification, which will be carried out by the CHIC data security provider on the instructions of the platform administrator:

- a second round of pseudonymisation, in which the initial code is substituted by a new code and the link between them held by a trusted third party);
- removal or offsetting of the most risky indirect identifier values in the patient dataset, including e.g. real diagnostic or treatment dates;
- further context-specific assessment of the residual risk of re-identification;

If, and only if, a CHIC end user discovers new information in the course of processing the data that could be of special clinical importance for a patient's individual treatment, it may contact the CDP/CHIC Platform Administrator; the latter could then with the aid of the trusted third party recover the original pseudonym supplied by the CHIC clinical user and alert the latter to the discovery so it can make use of the information as clinically appropriate.

Storage limitation

Storage of personal data is allowed by the Data Protection Directive 95/46/EC. But when the purpose of processing is achieved and the data are not required any more for that particular purpose, personal data must be rendered anonymous or must be destroyed. Most national laws allow personal data to be stored for a longer term, provided that this is in order to use the data exclusively to carry out scientific research or statistics. Nevertheless, some national laws impose supplementary conditions or formalities in order to allow longer storage.

Technical and organisational measures

Organisational measures, together with technical measures, must ensure an appropriate level of security of the data processing, taking into account the state of the art and the costs of their implementation relative to the risks inherent in the processing and the nature of the data to be protected. Appropriate organisational measures shall be taken by the controller against accidental loss, destruction or alteration of, or damage to, personal data and against unauthorised or unlawful processing of personal data in particular where the processing involves the transmission of data over a network, and against all other unlawful forms of processing. An indicative list of appropriate organisational measures that the CHIC platform administrator will take to ensure the confidentiality, integrity and accuracy of the data within the platform infrastructure appears in Annex B:

Annex B

Technical and organisational measures applying to data processed in the CHIC platform

(Version 2.0, September 2016)

The CDP/CHIC platform administrator will take appropriate technical and organisational measures to protect the data within the platform against misuse and loss, in accordance with European data protection rules, including all necessary and reasonable precautions:

- to prevent unauthorised persons from gaining access to data processing systems with which the data are processed or used (physical access control),
- to prevent data processing systems from being used without authorisation (denial of use control),
- to ensure that persons entitled to use a data processing system can gain access only to the data to which they have a right of access, and that the data cannot be read, copied, modified or removed without authorisation in the course of processing or use and after storage (data access control),
- to ensure, including through use of secure encryption, that the data cannot be read, copied, modified or removed without authorisation during electronic transmission, transport or storage and that it is possible to examine and establish to which bodies the transfer of personal data by means of data transmission facilities is envisaged (data transmission control),
- to ensure that it is possible retrospectively to examine and establish whether and by whom the data have been inputted into data processing systems, modified or removed (input control),
- to ensure that the data being processed on commission are processed solely in accordance with the directions of the controller (contractual control),
- to ensure that the data are protected against accidental destruction or loss (availability control),
- to ensure that data transferred and/or held for the different purposes is processed separately (separation rule). This means for example that the patient data used by clinicians to validate the CHIC models must be strictly separated from the data used by modelers to build or adapt and refine the models,

Annex C - ACCESS AUTHENTICATION FORM

(Version 2.0, September 2016)

I, the undersigned, (title) born on,
in..... and working on the project called CHIC on behalf of
..... (name of CHIC Platform Administrator or relevant CHIC modeler end
user) declare by this Access Authentication form, I am authorised to have access to the CHIC data.

I have read, understand and agree to observe the conditions as stated in the agreement as well as
Annexes A and B (the Explanatory Glossary and Technical and Organizational measures) which form
a part of this document (Version 2.0, March 2016).

I understand that two original copies of this agreement will be produced and will be kept by me and the
CHIC Platform Administrator respectively.

Signature of employee:

Date and Place:

Appendix 4 – CHIC End User Agreement

(Version 2.0, September 2016)

BETWEEN

the CHIC Platform Administrator

(address and country of establishment)

AND

("CHIC end user")

(address and country of establishment)

Individually referred to as a "Party" or collectively referred to as the "Parties".

Preamble

The Computational Horizons In Cancer (CHIC) project was a EU-financed FP7 project that ended on 31 March 2017, and which created an infrastructure for the development of integrative multiscale cancer models and hypermodel oncosimulators. These are continuing to be clinically adapted and partly validated in a process which involves sharing of clinical and genomic data of patients within the CHIC platform infrastructure. Each of the parties engaged in the validation recognises as a priority the imperative need to respect the fundamental interests and rights of patients, including the need to preserve the security and privacy of personal data being processed.

Therefore the infrastructure of CHIC is embedded in the CHIC Data Protection Framework, which guarantees compliance with current European data protection legislation, primarily by de facto anonymising the patient data. Due to the diverse participation of modellers and researchers in the project, it is of high importance to process patient data in compliance with all applicable laws and regulations, including without limitation, privacy and medical secrecy laws applicable to the activities of the parties. During the CHIC project, the CHIC platform infrastructure and security was administered by the CDP; after the end of the project, this role is being taken over by the CHIC platform administrator.

To fulfil the objectives of the model validation and adaptation, the patient data used by the modellers is subject to de-identification, using secure state of the art pseudonymisation/de-identification tool (eg, CATS), and stored in secure, access-controlled data repositories within the CHIC platform infrastructure. The CHIC platform administrator will transfer the original (data provider) code to an independent trusted third party, and the latter alone will retain the pseudonymisation key (cross table) needed to link the double-encrypted CHIC data set to the initial de-identified data sets provided by the clinical partners. The CHIC clinical users who transfer patient data to the CHIC infrastructure will ensure that before doing so they obtain fully informed consent from the patients concerned and relevant ethics and other regulatory body approvals data providers; in addition they will adhere fully to

the requirements of data protection, in implementing necessary technical and organisational safeguards in their institutions to protect the data.

This agreement is needed to state the conditions and obligations under which the CHIC end-users (modellers and tool developers) will process data within the said infrastructure.

Clause 1: Definitions

For the purposes of this agreement, the terms used in these clauses shall have the same meaning as attributed to them in the Explanatory Glossary in Annex A to this agreement.

Clause 2: Scope and responsibility

1. This Agreement sets out the terms and conditions for the CHIC modellers and component developers working to validate and adapt CHIC cancer models (end users) to access, use, and share patient data within the CHIC infrastructure.
2. The CHIC Platform administrator is responsible as data controller for the management of the CHIC infrastructure, while the CHIC end user is responsible for the data it accesses and uses from the infrastructure within its own organisation.

Clause 3: Obligations of the CHIC Platform Administrator

The **CHIC Platform Administrator** warrants and undertakes:

1. to grant to the CHIC end user a non-exclusive right to access and use the data in the CHIC data infrastructure (hereinafter the CHIC data) for the purposes of the end user's work in validating and adapting the cancer models, subject to the provisions of this agreement;
2. that it is entitled to grant access to the CHIC data to the end user as aforesaid;
3. to put in place procedures to ensure that prior to transfer to the CHIC infrastructure, CHIC data are collected and processed in accordance with the laws applicable to the data provider, including by entering into the 'CHIC Clinal User Agreement' with relevant data providers;
4. to ensure that the CHIC data made available to the end user has been securely de-identified, including by a double pseudonymisation process involving the use of an independent trusted third party to retain the pseudonymisation key (cross table).
5. to have in place appropriate technical and organisational measures in accordance with EU data protection law to protect patient data within the CHIC infrastructure against accidental or unlawful destruction or accidental loss, alteration, unauthorised disclosure or access, including, if it not in a position to provide the required measures through its own expertise, by entering the 'CHIC Data Security Agreement' with the CHIC data security provider. Where it provides its own security, it shall nonetheless make use of an independent trusted third party to double pseudonymise the data per subclause 3.4 above;
6. to ensure that each of its employees who has contact with the patient data is made aware of, and will be bound by, the terms of this Agreement, and such an employee will complete Annex C to this Agreement.

Clause 4: Obligations of the CHIC end user

The **CHIC end user** warrants and undertakes:

1. to process the CHIC data in compliance with applicable data protection regulation and the terms of this agreement; and where it cannot provide such compliance for whatever reasons, it agrees to inform promptly the CHIC Platform Administrator of its inability to comply, in which case the latter is entitled to suspend access to the data and/or terminate the contract;
2. to process the CHIC data only for the purposes of validating and adapting the CHIC cancer models;

3. that it has implemented and follows appropriate technical and organisational security measures to protect the CHIC data against misuse and loss (including without limitation the measures stated in Annex B to this agreement), in accordance with the requirements of relevant provisions of European data protection law, and in particular Article 17 of the Data Protection Directive 95/46/EC or Article 32 of the General Data Protection Regulation 2016/679 once this takes effect;
4. that it will ensure that where CHIC data is stored within its own organisation, such data is technically and organisationally separated from other data;
5. that it will retain the CHIC data within a secure database or network system at such standard as would be reasonably expected for the storage of sensitive/confidential data;
6. that it shall not attempt to identify any patient from the CHIC data either by external matching of the data or by any other means;
7. that it shall not disclose or publish the CHIC data to any third party, which for the avoidance of doubt includes any of its subcontractors or party with which it has an equivalent arrangement, without seeking and obtaining the specific written authorisation of the CHIC Platform Administrator;
8. that in the event of inadvertently identifying any patient, it shall notify the CHIC Platform Administrator immediately setting out (in reasonable detail) the circumstances by which this occurred. In such a case it further undertakes not to make any use of the identifying information for any purposes and to take all necessary steps to protect the interests of the patient including so far as possible restoring the de-identified status of the patient;
9. to ensure that each of its employees who has contact with the CHIC data is made aware of, and will be bound by, the terms of this Agreement, and such an employee will complete Annex C to this Agreement;
10. to inform the CHIC Platform Administrator immediately, should the CHIC data, while in the hands of the end user be threatened with seizure or confiscation through bankruptcy or settlement proceedings, or through any other circumstances including the actions of a third party.
11. that if it becomes aware that it is necessary or desirable, in the exceptional circumstances identified Annex A to this agreement, for the CHIC data to be re-linked to the data subject it shall contact the CHIC Platform Administrator only, so that the latter can initiate the re-identification process with the help of the Trusted Third Party that holds the key to link the de-identified data sets in respect of the subject concerned;
12. to deal promptly and properly with all inquiries from the CHIC Platform Administrator relating to its data processing and data security measures;
13. that upon reasonable request by the CHIC Platform Administrator, it will submit its data processing facilities, data files and documentation needed for reviewing, auditing and/or certifying by the CHIC Platform Administrator (or any independent or impartial inspection agents or auditors, selected by the same and not reasonably objected to by the CHIC end user) to ascertain compliance with the warranties and undertakings in these clauses, with reasonable notice and during regular business hours. The same obligations apply in case a supervisory authority demands auditing;
14. to provide the CHIC Platform Administrator with contact details of the person responsible for data protection in its organisation.

Clause 5: Cooperation with supervisory authorities

1. The CHIC Platform Administrator agrees to deposit a copy of this contract with the supervisory authority if it so requests or if such deposit is required under the applicable regulation.
2. The parties agree that the supervisory authority has the right to conduct an audit of the CHIC end user which has the same scope and is subject to the same conditions as would apply to an audit of the CHIC Platform Administrator under the applicable regulation.

Clause 6: Liability and indemnity

- .1. Each Party shall be liable to the other party for damages it causes by any breach of these clauses. The Parties agree that if one Party is held liable for a violation of the clauses committed by

the other Party, the latter will, to the extent to which it is liable, indemnify the first Party for any cost, charge, damages, expenses or loss it has incurred. Indemnification is contingent upon:

the Parties promptly notifying each other of a claim; and

each Party is given the possibility to cooperate in the defence and settlement of the claim.

2. The Parties agree that each Party shall be liable for patient's damages it caused by any negligent violation of data protection legislation or any analogous provisions of national or international law.

Clause 7: Penalty

1. The Parties agree that subject to the exception in clause 7.3 below, a Party in wilful or negligent breach of clause 3 or 4 of this agreement shall pay a penalty of 10.000 (ten thousand) EUR. The enforcement of this clause shall be subject to the finding of wilfulness or negligence by the court under Clause 9 below.

2. The penalty shall be paid to the Virtual Physiological Human Institute or relevant successor body to the same for the purpose of furthering research into the development and implementation of integrative multiscale cancer models.

3. In the event that the breach or series of breaches does not lead to the identification of any data subject, then provided that the Party in breach timeously corrects the breach in accordance with the terms of clause 8.2 below, it shall escape the liability set out in this clause.

4. The above provision shall be without prejudice to the Parties' right to terminate the contract, to seek compensation for damages or to enforce any claims under this agreement.

Clause 8: Termination and obligations of the parties after the termination

1. This agreement will terminate, if not otherwise superseded or amended by new provisions extending it, at the latest by: [Date].

2. In case of breach of clauses 3 or 4 by one of the Parties, the other Party is entitled to give written notice requiring the party in breach to be repair the breach within 72 hours, after which time if the breach remains outstanding it may terminate this agreement.

3. Without prejudice to the foregoing provisions, any party may terminate this agreement for good cause, giving the reason for such termination.

4. Each party shall inform the other party by prior written notice in case of termination of the agreement.

5. The parties agree that on the termination of the provision of data processing services, the CHIC end user shall, at the choice of the CHIC Platform Administrator, return all the CHIC data and the copies thereof to the CHIC Platform Administrator or shall destroy all the data and certify to the CHIC Platform Administrator that it has done so, unless legislation prevents the CHIC end user from so doing. In all cases, the CHIC end user warrants that it will continue to guarantee the confidentiality of the data and will no longer actively process the data.

Clause 9: Governing law and Jurisdiction, miscellaneous

1. This agreement shall be governed by Belgian Law. The courts of Brussels/Belgium shall have exclusive jurisdiction. This shall also apply to disputes on the validity of this clause.

2. Changes and amendments to this agreement shall require written agreement signed by the parties and an explicit statement that they represent a change or amendment to these conditions. The same applies to the waiving of this formal requirement.

3. If any provision of this agreement shall be entirely or partly invalid or unenforceable, this shall not affect the validity and enforceability of any other provision. An invalid or unenforceable provision shall be regarded as replaced by such a valid and enforceable provision that as closely as possible

reflects the privacy/security and/or economic purpose that the Parties hereto had purposed with the invalid or unenforceable provision.

4. Each person signing below and each Party on whose behalf such person executes this agreement warrants that he/she, as the case may be, has the authority and the legal capacity to enter into this contractual agreement and perform the obligation herein.

5. This agreement will enter into force on the effective date, i.e. the date of the last binding signature to this agreement.

Made in two signed copies, each party having received its own signed copy.

(Place, Date)

(Auth'd Rep. of CHIC Platform Administrator)

(Place, Date)

(Auth'd Rep of CHIC end user)

Annexes:

- A. Explanatory Glossary
- B. Technical and organisational measures
- C. Access authentication form

[Identical to Annexes of CHIC Clinical User Agreement (see Appendix 3) and omitted for reasons of space.]

Appendix 5 – CHIC Data Security Agreement

(Version 2.0, September 2016)

BETWEEN

the CHIC Platform Administrator

(address and country of establishment)

AND

("CHIC Data Security Provider")

(address and country of establishment)

Individually referred to as a "Party" or collectively referred to as the "Parties".

Preamble

The Computational Horizons In Cancer (CHIC) project was a EU-financed FP7 project that ended on 31 March 2017, and which created an infrastructure for the development of integrative multiscale cancer models and hypermodel oncosimulators. These are continuing to be clinically adapted and partly validated in a process which involves sharing of clinical and genomic data of patients within the CHIC platform infrastructure. Each of the parties engaged in the validation recognises as a priority the imperative need to respect the fundamental interests and rights of patients, including the need to preserve the security and privacy of personal data being processed.

Therefore the infrastructure of CHIC is embedded in the CHIC Data Protection Framework, which guarantees compliance with current European data protection legislation, primarily by de facto anonymising the patient data. Due to the diverse participation of modellers and researchers in the project, it is of high importance to process patient data in compliance with all applicable laws and regulations, including without limitation, privacy and medical secrecy laws applicable to the activities of the parties. During the CHIC project, the CHIC platform infrastructure and security was administered by the CDP; after the end of the project, this role is being taken over by the CHIC platform administrator. In order to fulfil its role as required in relation to data security provision assist the CHIC platform administrator may enter into a contract with a specialist data security provider.

To fulfil the objectives of the model validation and adaptation, the patient data used by the modellers is subject to de-identification, using secure state of the art pseudonymisation/de-identification tool (eg, CATS), and stored in secure, access-controlled data repositories within the CHIC platform infrastructure. The CHIC platform administrator will transfer the original (data provider) code to an independent trusted third party (which may be the data security provider), and the latter alone will retain the pseudonymisation key (cross table) needed to link the double-encrypted CHIC data set to the initial de-identified data sets provided by the clinical partners. The CHIC clinical users who transfer patient data to the CHIC infrastructure will ensure that before doing so they obtain fully informed consent from the patients concerned and relevant ethics and other regulatory body approvals data

providers; in addition they will adhere fully to the requirements of data protection, in implementing necessary technical and organisational safeguards in their institutions to protect the data.

This agreement is needed to state the conditions and obligations under which the CHIC Data Security Provider shall fulfil its function referred to above.

Clause 1: Definitions

For the purposes of this agreement, the terms used in these clauses shall have the same meaning as attributed to them in the Explanatory Glossary in Annex A to this agreement.

Clause 2: Scope and responsibility

1. This agreement sets out the terms and conditions under which the Data Security Provider shall provide and implement legally required technical data security measures in order to protect patient data in the CHIC infrastructure. It also includes the terms on which the Data Security Provider may securely retain the cross table / pseudonymisation key (hereafter “the key”) and assist the CHIC Platform Administrator in allowing a data provider to re-link data to a relevant patient subject to the feedback procedure described below.

2. The CHIC Platform Administrator is responsible as data controller for the management of the CHIC infrastructure, while the CHIC Data Security Provider is responsible for the technical data security to protect the infrastructure and data within it and the retention and security of the key as aforesaid.

Clause 3: Obligations of the CHIC Platform Administrator

The **CHIC Platform Administrator** warrants and undertakes:

1. to grant to CHIC Data Security Provider access to the CHIC infrastructure to the extent reasonably required for it to carry out its data security obligations under this agreement;
2. to grant to the CHIC Data Security Provider in its role as a trusted third party a right to retain the key, in order that it may fulfil its responsibilities as set out in clause 2 of this agreement;
3. to put in place procedures, including in particular by entering into a contractually binding “CHIC Clinical User Agreement” with each clinical data provider, to ensure that, prior to upload to the CHIC infrastructure, all patient data may be lawfully provided by the same and has been subjected to a preliminary pseudonymisation process;
4. to remunerate the CHIC Data Security Provider for its services under this agreement according to such terms as may be agreed between the Parties.

Clause 4: Obligations of the CHIC Data Security Provider

The **CHIC Data Security Provider** warrants and undertakes:

1. to implement state of the art technical and organisational security measures to protect the CHIC infrastructure and data within the infrastructure against misuse and loss (including without limitation the measures stated in Annex B to this agreement), in accordance with the requirements of relevant provisions of European data protection law, and in particular Article 17 of the Data Protection Directive 95/46/EC or Article 32 of the General Data Protection Regulation 2016/679 once this takes effect;
2. to retain, in its capacity of trusted third party, the key in a secure manner, protected by state of the art access security, and to deploy the said key only in the strict circumstances set out in the present Clause;
3. that it shall not disclose the key to any third party, which for the avoidance of doubt includes any of its subcontractors or party with which it has an equivalent arrangement, without seeking and obtaining the specific written authorisation of the CHIC Platform Administrator;

4. that except as specifically provided for in subclause 4.6 below, it shall take no measures that may conduce by any means to the re-identification of the data within the CHIC infrastructure, including of patients who are subjects of the codes contained in the cross table making up the key;
5. that if it is contacted by any CHIC end user that, through analyzing the CHIC data, has acquired information of potential importance to the original patient subject, and which requests it to deploy the key to enable re-linking of the securely de-identified data to the patient in question, it shall refer the request to the CHIC Platform Administrator;
6. that only if so instructed by the CHIC Platform Administrator, it shall take the steps in this sub-clause to comply with the request: first it will ask the end user for the code attached to the data set following the second round of encryption; second it will deploy its key in order to discover the original code attached by data provider; third it will notify the CHIC Platform Administrator of the original code so that the latter can alert the CHIC clinical user where the patient is treated that there is new information pertaining to the patient for the responsible clinician to use the information as clinically appropriate;
7. to ensure that it processes any data hosted in the CHIC infrastructure only to the extent strictly required to fulfil its security obligations under this agreement; in such case it shall at all times maintain the strictest security and confidentiality in relation to the said data and ensure that it is permanently erased from and/or no longer accessible to its systems as soon as possible;
8. to ensure that each of its employees who has contact with the key and/or any data hosted in the CHIC infrastructure is made aware of, and will be bound by, the terms of this Agreement, and such an employee will complete Annex C to this Agreement; .
9. to deal promptly and properly with all inquiries from the CHIC Platform Administrator relating to its data security measures and faithfully follow the instructions of the CHIC Platform Administrator.

Clause 5: Cooperation with supervisory authorities

1. The CHIC Platform Administrator agrees to deposit a copy of this agreement with the supervisory authority if it so requests or if such deposit is required under the applicable regulation.
2. The Parties agree that the supervisory authority has the right to conduct an audit of the CHIC Data Security Provider which has the same scope and is subject to the same conditions as would apply to an audit of the CHIC Platform Administrator under the applicable regulation.

Clause 6: Liability and indemnity

1. Each Party shall be liable to the other Party for damages it causes by any breach of these clauses. The Parties agree that if one Party is held liable for a violation of the clauses committed by the other Party, the latter will, to the extent to which it is liable, indemnify the first Party for any cost, charge, damages, expenses or loss it has incurred. Indemnification is contingent upon:

the Parties promptly notifying each other of a claim; and

each Party is given the possibility to cooperate in the defence and settlement of the claim.

2. The Parties agree that each Party shall be liable for patient's damages it caused by any negligent violation of data protection legislation or any analogous provisions of national or international law.

Clause 7: Termination and obligations of the parties after the termination

1. This agreement will terminate, if not otherwise superseded or amended by new provisions extending it, at the latest by:[Date].
2. In case of breach of clauses 3 or 4 by one of the Parties, the other Party is entitled to give written notice requiring the Party in breach to be repair the breach within 72 hours, after which time if the breach remains outstanding it may terminate this agreement.
3. Without prejudice to the foregoing provisions, each Party may terminate this agreement for good cause, giving the reason for such termination.

4. Each Party shall inform the other party by prior written notice in case of termination of the agreement.

5. The Parties agree that on the termination of the provision of data processing services, the CHIC Data Security Provider shall destroy the key or any other data from the CHIC infrastructure and certify to the CHIC Platform Administrator that it has done so, unless legislation imposed upon the CHIC Data Security Provider prevents it from so doing. In all cases, the CHIC Data Security Provider warrants that it will continue to guarantee the security and confidentiality of the key or any other data.

Clause 8: Governing law and Jurisdiction, miscellaneous

1. This agreement shall be governed by Belgian Law. The courts of Brussels/Belgium shall have exclusive jurisdiction. This shall also apply to disputes on the validity of this clause.

2. Changes and amendments to this agreement shall require written agreement signed by the parties and an explicit statement that they represent a change or amendment to these conditions. The same applies to the waiving of this formal requirement.

3. If any provision of this agreement shall be entirely or partly invalid or unenforceable, this shall not affect the validity and enforceability of any other provision. An invalid or unenforceable provision shall be regarded as replaced by such a valid and enforceable provision that as closely as possible reflects the privacy/security and/or economic purpose that the Parties hereto had purposed with the invalid or unenforceable provision.

4. Each person signing below and each Party on whose behalf such person executes this agreement warrants that he/she, as the case may be, has the authority and the legal capacity to enter into this contractual agreement and perform the obligation herein.

5. This agreement will enter into force on the effective date, i.e. the date of the last binding signature to this agreement.

Made in two signed copies, each Party having received its own signed copy.

(Place, Date)

(Auth'd Rep. of CHIC Platform Administrator)

(Place, Date)

(Auth'd Rep of CHIC Data Security Provider)

Annexes:

- A. Explanatory Glossary
- B. Technical and organisational measures
- C. Access authentication form

[Identical to Annexes of CHIC Clinical User Agreement (see Appendix 3) and omitted for reasons of space.]

Appendix 6 – CHIC software licensing report

N	Component Party	Dependencies Licenses	CHIC interaction	License Compatibility	Component license
1	ObTiMA (data management system) USAAR	<p>Calling object code: -----</p> <p>Apache License 2.0 Berkeley Software Distribution (BSD) 3-Clause License Common Development and Distribution License (CDDL) 1.1 Common Public License (CPL) 1.0 Eclipse Public License (EPL) 1.0 GNU Lesser General Public License (LGPL) 2.1</p> <p>Other License Types</p> <ul style="list-style-type: none"> - ANTLR 2¹⁵² - Bouncy Castle/MIT License¹⁵³ - dom4j/BSD¹⁵⁴ 	N/A	<p>BSD 3-Clause license, MIT license are compatible with Apache v2¹⁵⁹ and GPL¹⁶⁰. Apache v2 allows codes under CDDL, CPL and EPL in binaries upon appropriate labeling. Apache v2 does not allow LGPLv2 in Apache products¹⁶¹.</p> <p>LGPL may be combined into larger works with Apache v2, CDDL, CPL and EPL, provided the CDDL, CPL and EPL license terms are observed and license allows modification and decompilation for the customer's own use. All licenses allow development and redistribution of larger works, provided the initial codes stay under the respective licenses and the license terms for them are met by the redistribution,</p>	<p>No binding license.</p> <p>Licensing under GPL would create incompatibility problems with EPL, CDDL, CPL¹⁶². If packaged with LGPL dependencies, licensing under Apache might create incompatibility issues with LGPL, because Apache does not tolerate LGPL in Apache products¹⁶³. If distributed with LGPL dependencies Component license must allow modification and reverse engineering for the customer's own use (para 6 LGPLv.2.1).</p> <p>Distribution in object code (on commercial basis) and source code possible.</p> <p>Suggested open source licenses: BSD 3-Clause License, MIT License.</p> <p>Redistribution requirements (if packaged with dependencies):</p> <p>Codes of dependencies must stay under original licenses. Please keep copyright, disclaimers and license notices intact. Provide instructions to the user where/how to get the CDDL, CPL, EPL</p>

¹⁵² "We encourage users to develop software with ANTLR. However, we do ask that credit is given to us for developing ANTLR. By "credit", we mean that if you use ANTLR or incorporate any source code into one of your programs (commercial product, research project, or otherwise) that you acknowledge this fact somewhere in the documentation, research report, etc... If you like ANTLR and have developed a nice tool with the output, please mention that you developed it using ANTLR. In addition, we ask that the headers remain intact in our source code. As long as these guidelines are kept, we expect to continue enhancing this system and expect to make other tools available as they are completed". See: <http://www.antlr2.org/license.html>.

¹⁵³ <https://www.bouncycastle.org/licence.html>.

¹⁵⁴ <http://dom4j.sourceforge.net/dom4j-1.6.1/license.html>



		<ul style="list-style-type: none"> - jsoup/MIT License¹⁵⁵ - jTidy¹⁵⁶ - Sesame/BSD 3-Clause¹⁵⁷ - SLF4J/MIT License¹⁵⁸ 			<p>source codes.</p> <p>BSD 3-Clause: reproduce copyright notice, license terms and disclaimers</p> <p>CDDL 1.1: make available CDDL source code, provide license along with the code; keep intact copyright and other notices.</p> <p>EPL 1.0: make available EPL source code; provide EPL license, keep intact copyright notices</p> <p>CPL 1.0: inform how to get source code from the source, reproduce copyright notices, license terms, disclaimers.</p> <p>LGPL 2.1: credit LGPL library, provide LGPL license for the library, inform how to get the source for the library. Please see para 6 LGPL 2.1 for more details¹⁶⁴.</p> <p>Apache v2: provide license, retain copyright and disclaimer notices.</p> <p>ANTLR: acknowledge ANTLR, keep intact headers in source code.</p> <p>Dom4j: reproduce copyright notice, license terms and disclaimers; give credit to the DOM4J Project</p>
--	--	--	--	--	--

¹⁵⁹ <http://www.apache.org/legal/resolved.html>

¹⁶⁰ <https://www.gnu.org/licenses/license-list.en.html>

¹⁶¹ <https://www.gnu.org/licenses/license-list.en.html>

¹⁶² <https://www.gnu.org/licenses/license-list.en.html>

¹⁶³ <http://www.apache.org/legal/resolved.html>

¹⁵⁵ <https://jsoup.org/license>

¹⁵⁶ <http://jtidy.sourceforge.net/license.html>

¹⁵⁷ <https://bitbucket.org/openrdf/sesame/src/master/core/LICENSE.txt>

¹⁵⁸ <http://www.slf4j.org/license.html>

¹⁶⁴ <https://opensource.org/licenses/LGPL-2.1>

					<p>- http://dom4j.sourceforge.net</p> <p>MIT: include copyright notice and license terms</p> <p>jTidy: keep intact copyright notice.</p>
2	<p>Robust Interactive Multi-label</p> <p>(software)</p> <p>USAAR</p>	<p>Calls object code:</p> <p>-----</p> <p>Apache License 2.0</p> <p>BSD 3-Clause</p> <p>GNU Lesser General Public License (LGPL) 2.1</p>	N/A	<p>BSD 3-Clause is compatible with Apache v2¹⁶⁵ and LGPL¹⁶⁶.</p> <p>LGPL v 2.1 may be combined with Apache v2 (para 6 LGPL v2.1).</p> <p>LGPL v2.1 may not be included into Apache software¹⁶⁷.</p>	<p>No binding license.</p> <p>If packaged with LGPL dependencies, licensing under Apache might create incompatibility issues with LGPL, because Apache does not tolerate LGPL in Apache products¹⁶⁸. License must allow modification and reverse engineering for the customer's own use (para 6 LGPLv.2.1).</p> <p>Distribution in object code and source code possible. Licensing under GPL would require release of the source code and non-commercial licensing only</p> <p>Suggested open source licenses: copyleft GPLv3¹⁶⁹, permissive licenses: BSD 3-Clause, Apache v2 (if packaged without LGPL)</p> <p>Redistribution requirements (if packaged with dependencies):</p> <p>Codes of dependencies must stay under original licenses. Please keep copyright, disclaimers and license notices intact.</p> <p>LGPL 2.1: credit LGPL library, provide LGPL license for the library, inform how to get the</p>

¹⁶⁵ <http://www.apache.org/legal/resolved.html>

¹⁶⁶ <https://www.gnu.org/licenses/license-list.en.html>

¹⁶⁷ <http://www.apache.org/legal/resolved.html>

¹⁶⁸ <http://www.apache.org/legal/resolved.html>

¹⁶⁹ <https://opensource.org/licenses/GPL-3.0>

					<p>source for the library. Please see para 6 LGPL 2.1 for more details¹⁷⁰.</p> <p>Apache v2: provide license, retain copyright and disclaimer notices.</p> <p>BSD 3-Clause: reproduce copyright notice, license terms and disclaimers.</p>
3	Automatic brain tumor segmentation with a fast Mumford-Shah algorithm (software). USAAR	<p>Calls the object code: -----</p> <p>Apache License 2.0: Berkeley Software Distribution (BSD) 3</p>	N/A	Apache v2 and BSD 3-Clause are compatible with each other ¹⁷¹ .	<p>No binding license.</p> <p>Distribution in object code on proprietary basis and source code possible.</p> <p>Suggested open source license: Apache v2.</p> <p>Redistribution requirements (if packaged with dependencies):</p> <p>Codes of dependencies must stay under original licenses. Please keep copyright, disclaimers and license notices intact.</p> <p>Apache v2: provide license, retain copyright and disclaimer notices.</p> <p>BSD 3-Clause: reproduce copyright notice, license terms and disclaimers.</p>
4	Clinical Data Repository (repository architecture) UBERN	<p>Calls the object code, HTTP calls: -----</p> <p>Apache v2: - ASP.NET - Entity Framework - SimpleITK¹⁷²</p>	<p>HTTPcalls -----</p> <p>RICORDO - LOLS (HTTP calls) - Rdfstore (HTTP calls)</p>	<p>Apache v2 is compatible with Microsoft Public License, BSD 3-Clause License and MIT License. SIL OFL may be included into Apache software if inclusion is properly labeled¹⁸³.</p> <p>Communication via HTTP calls is</p>	<p>No binding license.</p> <p>Distribution in object code on proprietary basis and source code possible.</p> <p>Suggested open source license: Apache v2¹⁸⁴.</p> <p>Redistribution requirements(if packaged with</p>

¹⁷⁰ <https://opensource.org/licenses/LGPL-2.1>

¹⁷¹ <http://www.apache.org/legal/resolved.html>

¹⁷² <http://www.simpleitk.org>

	<ul style="list-style-type: none"> - ReCaptcha¹⁷³ - Fuseki¹⁷⁴ - Microsoft Public License (MPL): - Fellow Oak DICOM for .NET MIT License: - dotNetRDF - Newtonsoft Json¹⁷⁵ - VDS.Common¹⁷⁶ - jQuery¹⁷⁷ - Bootstrap¹⁷⁸ - Other Licenses: - HDF5DotNet (similar to BSD 3-Clause)¹⁷⁹ - Statismo/BSD 3-Clause License¹⁸⁰ - FontAwesome: Fonts: SIL OFL 1.1, code license: MIT License¹⁸¹ 	Custodix <ul style="list-style-type: none"> - Authentication service (HTTP calls) - Auditing service (HTTP calls) 	communication between separate programs and does not produce licensing implications.	dependencies): Codes of dependencies must stay under original licenses. Please keep copyright, disclaimers and license notices intact. HDF5DotNet: reproduce copyright notice, license conditions, and disclaimer; credit to HDF Group and by the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign and credit the contributors. SIL OFL 1.1: retain copyright notices and license, font software must stay under the OFL 1.1. BSD 3-Clause: reproduce copyright notice, license terms and disclaimers. MIT: include copyright notice and license terms
--	---	---	--	---

¹⁸³ <http://www.apache.org/legal/resolved.html>

¹⁸⁴ <https://opensource.org/licenses/Apache-2.0>

¹⁷³ <https://developers.google.com/recaptcha/>

¹⁷⁴ http://jena.apache.org/documentation/serving_data/

¹⁷⁵ <https://github.com/JamesNK/Newtonsoft.Json/blob/master/LICENSE.md>

¹⁷⁶ <https://www.nuget.org/packages/VDS.Common/>

¹⁷⁷ <https://jquery.org/license/>

¹⁷⁸ <http://www.getbootstrap.com>

¹⁷⁹ <http://www.hdfgroup.org/HDF5/doc/Copyright.html>

¹⁸⁰ <https://github.com/statismo/statismo>

¹⁸¹ <http://fontawesome.github.io/Font-Awesome/license>

		- Google Web Fonts: SIL OFL 1.1, Apache v2 ¹⁸²			
5	Bratumia (software) UBERN	<p>Calls object code: -----</p> <p>Apache 2.0 license: -ITK¹⁸⁵</p> <p>Microsoft Research License Agreement: -Sherwood library¹⁸⁶</p>	<p>via Plugin interface -----</p> <p>Dr. Eye (communication via Plugin interface of Dr. Eye)</p>	<p>Apache license does not consider linking (or binding by name) to the interfaces as creating a derivative work (para 1 Apache v2).</p> <p>MSR-LA allows creation of derivative works from the source code and distribution of such works for non-commercial purposes only under MSR-LA (para 3 MSR-LA).</p> <p>It may be argued that Apache v2 and MSR-LA can be combined by linking into a larger work and distributed as a whole for non-commercial purposes, provided the terms of Apache and MSR-LA for the original codes are met.</p> <p>Communication with Dr.Eye via Plugin interface does not produce licensing implications for Bratumia.</p>	<p>Non-commercial licensing only. Proprietary licensing not allowed by MSR-LA, requires Commercial Microsoft license.</p> <p>Distribution in object code and source code possible.</p> <p>Suggested open source license: Apache v2 non-commercial.</p> <p>Redistribution requirements (if packaged with dependencies):</p> <p>Codes of dependencies must stay under original licenses. Please keep copyright, disclaimers and license notices intact.</p> <p>MSR-LA: keep copyright, disclaimer and other notices intact, use for non-commercial purposes only, use or distribution of Software or any derivative works in any form for commercial purposes is not allowed; Microsoft software stays and may be distributed under MSR-LA only.</p> <p>Apache v2: provide license, retain copyright and disclaimer notices.</p>
6	CRAF (software) FORTH	<p>Calls object code: -----</p> <p>Apache v2</p>	<p>HTTP calls -----</p> <p>CRAF makes HTTP</p>	<p>BSD 3-Clause License and MIT License¹⁹⁷ are compatible with Apache¹⁹⁷ and LGPL¹⁹⁸. Apache does not tolerate LGPL software in</p>	<p>No binding license.</p> <p>If packaged with LGPL'd XOM, licensing under Apache might create incompatibility issues with LGPL, because Apache does not tolerate LGPL in</p>

¹⁸² <https://www.google.com/fonts/attribution>
¹⁸⁵ <https://itk.org/ITK/project/license.html>
¹⁸⁶ <https://www.microsoft.com/en-us/download/details.aspx?id=52340>

	<p>MIT License:</p> <ul style="list-style-type: none"> - Lombok¹⁸⁷ - Angular JS¹⁸⁸ - Angular Material¹⁸⁹ - MomentJS¹⁹⁰ - ngStorage¹⁹¹ - angular-fontawesome¹⁹² - ng-lodash¹⁹³ - Angular UI Grid (ui-grid)¹⁹⁴ <p>BSD 3-Clause License:</p> <ul style="list-style-type: none"> - OWNER¹⁹⁵ <p>GNU LGPL 2.1:</p> <ul style="list-style-type: none"> - XOM¹⁹⁶ 	<p>calls to the following components:</p> <ul style="list-style-type: none"> - Model Repository - inSilico Trial Repository - VPH-Execution Framework - Security and Anon/tion services (e.g. Identity Provider, PIMS) 	<p>Apache products¹⁹⁹. Apache v2 is compatible with LGPL²⁰⁰.</p> <p>Communication via HTTP calls is communication between separate programs and does not produce licensing implications.</p>	<p>Apache products²⁰¹. If CRAF links to the LGPL XOM from the web or if XOM is distributed and installed separately, Apache is ok. The component license must allow modifications and reverse engineering for the customer's own use²⁰². The component may go under LGPL and GPLv3, because GPL tolerates Apache v2 in GPLv3 products²⁰³.</p> <p>Suggested open source licenses:</p> <ul style="list-style-type: none"> - weak copyleft: LGPL 2.1²⁰⁴, LGPL 3.0²⁰⁵, - strong copyleft: GPLv3²⁰⁶ (GPLv2 is incompatible with Apache²⁰⁷), - lax permissive licenses: Apache v2 (if packaged without LGPL XOM), BSD 3 Clause License²⁰⁸.
--	--	--	--	--

¹⁹⁷ <http://www.apache.org/legal/resolved.html#category-a>

¹⁹⁸ <https://www.gnu.org/licenses/license-list.en.html#GPLCompatibleLicenses>

¹⁸⁷ <https://projectlombok.org/>

¹⁸⁸ <https://github.com/angular/angular.js>

¹⁸⁹ <https://material.angularjs.org>

¹⁹⁰ <https://github.com/moment/moment>

¹⁹¹ <https://github.com/gsklee/ngStorage>

¹⁹² <https://github.com/picardy/angular-fontawesome>

¹⁹³ <https://github.com/rockabox/ng-lodash>

¹⁹⁴ <https://github.com/angular-ui/ui-grid/>

¹⁹⁵ <https://github.com/lviggiano/owner>

¹⁹⁶ <http://www.xom.nu/>

¹⁹⁹ <http://www.apache.org/legal/resolved.html#category-x>

²⁰⁰ <https://www.gnu.org/licenses/license-list.en.html#GPLCompatibleLicenses>

²⁰¹ <http://www.apache.org/legal/resolved.html>

²⁰² para 6 LGPL.

					<p>Distribution in object code and source code possible. Licensing under GPLv3 would require release of the source code (also for derivatives) and non-commercial licensing only.</p> <p>Redistribution requirements (if packaged with dependencies):</p> <p>Codes of dependencies must stay under original licenses. Please keep copyright, disclaimers and license notices intact.</p> <p>BSD 3-Clause: reproduce copyright notice, license terms and disclaimers</p> <p>Apache v2: provide license, retain copyright and disclaimer notices.</p> <p>MIT: include copyright notice and license terms</p> <p>LGPL 2.1: credit LGPL library, provide LGPL license for the library, inform how to get the source for the library. Please see para 6 LGPL 2.1 for more details²⁰⁹.</p>
7	Lung cancer oncosimulator (mechanistic) (model)	<p>Calls object code:</p> <p>-----</p> <p>LGPL v3:</p> <p>- Muscle 2²¹⁰</p>	<p>Command line arguments, Model linking via Muscle library (LGPLv3):</p> <p>-----</p>	<p>Linking to LGPLv3 MUSCLE and interaction with the CHIC components/models via MUSCLE API does not produce licensing implications, because LGPLv3 license, Section 4 allows “combining</p>	<p>No binding license.</p> <p>May be distributed open source and in object code on proprietary basis. If packaged with LGPL MUSCLE, licensing under Apache might create incompatibility issues with LGPL, because Apache</p>

²⁰³ <https://www.gnu.org/licenses/license-list.en.html#GPLCompatibleLicenses>

²⁰⁴ <https://opensource.org/licenses/LGPL-2.1>

²⁰⁵ <https://opensource.org/licenses/LGPL-3.0>

²⁰⁶ <https://opensource.org/licenses/GPL-3.0>

²⁰⁷ <https://www.gnu.org/licenses/license-list.en.html#GPLCompatibleLicenses>

²⁰⁸ <https://opensource.org/licenses/BSD-3-Clause>

²⁰⁹ <https://opensource.org/licenses/LGPL-2.1>

²¹⁰ <http://apps.man.poznan.pl/trac/muscle>

	ICCS-NTUA Eleni Kolokotroni, Georgios Stamatakis, Dimitra Dionysiou		a)Lung cancer multimodeler hypermodel, molecular model (UPENN): static/sequential communication through command line argument, b)preprocessing tool (FORTH): static/sequential communication through command line argument, c)Biomechanics simulator (UBERN): dynamic/iterative communication through MUSCLE library, d)vasculature model (OXFORD): dynamic/iterative communication through MUSCLE library, e)metabolic model(FORTH): dynamic/iterative communication through MUSCLE library	or linking an Application with the Library” and to “convey a Combined Work under terms of your choice, that, taken together, effectively do not restrict modification of the portions of the Library contained in the Combined Work and reverse engineering for debugging such modifications” ²¹¹ Communication via command line arguments is a method of communication usually used between two separate programs ²¹² . When components interact in this way their licenses are independent from each other.	does not tolerate LGPL in Apache products ²¹³ . Suggested open source license: LGPL v3, Apache v2²¹⁴ (if packaged without MUSCLE) Redistribution requirements (if packaged with dependencies): Codes of dependencies must stay under original licenses. Please keep copyright, disclaimers and license notices intact. LGPL v3: credit LGPL library, provide LGPL license for the library, inform how to get the source for the library. Please see para 4 LGPL v3 for more details ²¹⁵ .
8	Glioblastoma	Calls the object code:	Single machine learning	LGPL allows development and distribution of linking software under	No binding license.

²¹¹ <https://opensource.org/licenses/LGPL-3.0>

²¹² <https://www.gnu.org/licenses/gpl-faq#MereAggregation>

²¹³ <http://www.apache.org/legal/resolved.html>

²¹⁴ <https://opensource.org/licenses/Apache-2.0>

²¹⁵ <https://opensource.org/licenses/LGPL-3.0>

	oncosimulator (model): ICCS-NTUA Markos Antonopoulos, Georgios Stamatakos, Dimitra Dionysiou	----- LGPL v3: - Muscle 2 Matlab runtime ²¹⁶	model	any license terms, which allow modification and reverse engineering for the customer's own use (para 4 LGPL v3). MCR is a non-copyleft license and does not place license restriction on licensing of applications, which run in Matlab runtime environment (para 1 MCR).	May be distributed open source and in object code on proprietary basis. Suggested open source licenses: LGPLv3, Apache v2²¹⁷ (if packaged without MUSCLE) Redistribution requirements (if packaged with dependencies): Codes of dependencies must stay under original licenses. Please keep copyright, disclaimers and license notices intact. LGPL v3: credit LGPL library, provide LGPL license for the library, inform how to get the source for the library. Please see para 4 LGPL v3 for more details ²¹⁸ .
9	Model/Tool Repository (Repository architecture) ICCS-NTUA Nikolaos Touser , Georgios Stamatakos, Dimitra Dionysiou	GPLv2: - MySQL community edition: Unix sockets BSD 3-Clause License: -Django Rest Framework: calling object code dm.xmlsec.binding: calling object code Django: calling object code	http calls/AMQP protocol/interfaces ----- The model/tool repository communicates with the editor, the CRAF, the VPH-HF and the CHIC triplestore through http calls or through the AMQP protocol. It is also integrated with the CHIC security framework by using	Provided the repository communicates with GPLv2 MySQL via sockets and MySQL is installed separately (not included into the repository distribution), GPLv2 copyleft shall not affect licensing of the repository. BSD, MIT and PSF licenses are compatible with GPL ²²¹ and Apache v2 ²²² . Apache v2 is incompatible with GPLv2, but may be included in GPLv3 software ²²³ .	No binding license. It may be distributed in object code on proprietary basis and open source. Suggested open source license: LGPLv3, -Apache v2²²⁶ Redistribution requirements (if packaged with dependencies): Codes of dependencies must stay under original licenses. Please keep copyright, disclaimers and license notices intact. BSD 3-Clause: reproduce copyright notice,

²¹⁶ "1. LICENSE GRANT. Subject to the restrictions below, The MathWorks, Inc. ("MathWorks") hereby grants to you, whether you are an individual or an entity, a license to install and use the MATLAB Compiler Runtime Libraries ("MCR"), solely and expressly for the purpose of running software created with the MATLAB Compiler (the "Application Software"), and for no other purpose. This license is personal, nonexclusive, and nontransferable". See MCR: http://www.tut.fi/cytospectre/cytospectre/MCR_license.txt

²¹⁷ <https://opensource.org/licenses/Apache-2.0>

²¹⁸ <https://opensource.org/licenses/LGPL-3.0>

		<p>Apache v2:</p> <ul style="list-style-type: none"> - djangosaml2²¹⁹: calling object code <p>PSF 2.7:</p> <p>Python 2.7²²⁰: programming language</p> <p>MIT License:</p> <ul style="list-style-type: none"> - XML security library: command line arguments - jQuery library: calling object code <p>Bootstrap Framework: calling object code</p>	<p>standardized interfaces (OASIS SAML standard and WS-* standard)</p>	<p>In terms of FSF “<i>pipes, sockets and command-line arguments are communication mechanisms normally used between two separate programs. So when they are used for communication, the modules normally are separate programs.</i>”²²⁴</p> <p>If GPLv2 code does not run against the repository, GPL copyleft shall not affect licensing of the repository.</p> <p>Communication via HTTP calls (sockets) is communication between separate programs and does not produce licensing implications²²⁵.</p>	<p>license terms and disclaimers</p> <p>MIT: include copyright notice and license terms</p> <p>Apache v2: provide license, retain copyright and disclaimer notices.</p> <p>GPLv2: publish on each copy an appropriate copyright notice and disclaimer of warranty; keep intact all the notices that refer to this License and to the absence of any warranty; and give any other recipients of the Program a copy of this License along with the Program.</p> <p>PSF: PSF's License Agreement and PSF's notice of copyright, i.e., "Copyright (c)</p> <p>2001, 2002, 2003, 2004, 2005, 2006 Python Software Foundation; All Rights Reserved" are retained in Python alone or in any derivative version.</p>
10	<p>In Silico Trial Repository</p> <p>(repository architecture)</p> <p>ICCS-NTUA</p>	See N 9.	<p>Http calls, interfaces</p> <p>-----</p> <p>The in silico trial repository communicates with the editor, the CRAF and the VPH-HF through</p>	See N 9	See N 9

²²¹ <https://www.gnu.org/licenses/license-list.en.html#ModifiedBSD>

²²² <http://www.apache.org/legal/resolved.html#category-a>

²²³ <http://www.apache.org/licenses/GPL-compatibility.html>

²²⁶ <https://opensource.org/licenses/Apache-2.0>

²¹⁹ <https://pypi.python.org/pypi/djangosaml2/>

²²⁰ <https://www.python.org/download/releases/2.7/license/>

²²⁴ <https://www.gnu.org/licenses/old-licenses/gpl-2.0-faq.en.html#MereAggregation>

²²⁵ <https://www.gnu.org/licenses/gpl-faq#MereAggregation>

	Nikolaos Touser, Georgios Stamatakos, Dimitra Dionysiou		http calls. It is also integrated with the CHIC security framework by using standardized interfaces (OASIS SAML standard and WS-* standard)		
11	NSCLC machine learning based response predictor (model) ICCS-NTUA Eleni Kolokotroni, Georgios Stamatakos, Dimitra Dionysiou	Matlab runtime environment: MCR ²²⁷	N/A	The MATLAB Runtime is a standalone set of shared libraries that enables the execution of compiled MATLAB applications or components on computers that do not have MATLAB installed. MCR does not affect licensing of the model, running on MATLAB compiler ²²⁸	No binding license. May be distributed in object code on proprietary basis and open source. Suggested open source license: Apache v2²²⁹.
12	Nephroblastoma oncosimulator (model) ICCS-NTUA	Calls object code: ----- Muscle 2: LGPL v3 ²³⁰ library (b5)	Command line arguments, MUSCLE API ----- a)Nephroblastoma multimodeler	Linking to LGPLv3 MUSCLE and interaction with the CHIC components/models via MUSCLE API does not produce licensing implications, because LGPLv3 license, Section 4 allows “ <i>combining or linking an Application with the</i>	No binding license. May be distributed open source and in object code on proprietary basis. If packaged with LGPL MUSCLE, licensing under Apache might create incompatibility issues with LGPL, because Apache does not tolerate LGPL in Apache products ²³³ .

²²⁷ <http://www.mathworks.com/products/compiler/mcr/>

²²⁸ “1. LICENSE GRANT. Subject to the restrictions below, The MathWorks, Inc. ("MathWorks") hereby grants to you, whether you are an individual or an entity, a license to install and use the MATLAB Compiler Runtime Libraries ("MCR"), solely and expressly for the purpose of running software created with the MATLAB Compiler (the "Application Software"), and for no other purpose. This license is personal, nonexclusive, and nontransferable.” See para 1 MCR.

²²⁹ <https://opensource.org/licenses/Apache-2.0>

²³⁰ <http://apps.man.poznan.pl/trac/muscle>

	Eleni Georgiadi, Georgios Stamatakis, Dimitra Dionysiou		<p>hypermodel, Molecular hypomodel (UPENN):static/sequential</p> <p>b) Preprocessing tool (FORTH): static/sequential</p> <p>c) Metabolic hypomodel (FORTH):Dynamic/iterative communication through MUSCLE library</p> <p>d) Vasculature hypomodel (UOXF): Dynamic/iterative communication through MUSCLE</p> <p>e) Biomechanical hypomodel: Dynamic/iterative communication through MUSCLE library</p>	<p><i>Library” and to “convey a Combined Work under terms of your choice, that, taken together, effectively do not restrict modification of the portions of the Library contained in the Combined Work and reverse engineering for debugging such modifications”²³¹</i></p> <p>Communication via command line arguments is a method of communication usually used between two separate programs²³². When components interact in this way their licenses are independent from each other.</p>	<p>Suggested open source licenses: LGPL v3, Apache v2²³⁴ (if packaged without LGPL MUSCLE).</p> <p>Redistribution requirements (if packaged with dependencies):</p> <p>Codes of dependencies must stay under original licenses. Please keep copyright, disclaimers and license notices intact.</p> <p>LGPL v3: credit LGPL library, provide LGPL license for the library, inform how to get the source for the library. Please see para 4 LGPL v3 for more details²³⁵.</p>
13	Security software component CUSTODIX	Dynamic and static linking: ----- CDDL 1.0 CDDL 2.0	N/A	The dependencies licenses are permissive copyright licenses (MIT, BSD, Apache) or licenses with weak copyleft (LGPL, CDDL, CPL, EPL). These licenses are compatible with each other and do not consider linking (static or dynamic) as	<p>No binding license.</p> <p>Licensing under GPL would create incompatibility problems with CDDL, EPL, CPL²⁴⁰. Component may go under proprietary license, provided the license terms for distribution of dependencies are observed.</p>

²³³ <http://www.apache.org/legal/resolved.html>

²³¹ <https://opensource.org/licenses/LGPL-3.0>

²³² <https://www.gnu.org/licenses/gpl-faq#MereAggregation>

²³⁴ <https://opensource.org/licenses/Apache-2.0>

²³⁵ <https://opensource.org/licenses/LGPL-3.0>

		<p>Apache License 2.0 Apache License 1.0 Apache License 1.1 BSD License</p> <p>BSD-2-Clause License BSD-3-Clause License Eclipse Public License 1.0 (EPL 1.0)</p> <p>Bouncy Castle License (=MIT License)</p> <p>France Telecom Copyright License²³⁶ (=BSD 3-Clause)</p> <p>LGPL 2.1 LGPL 3.0</p> <p>Mozilla Public License 2.0 (MPL 2.0) Jaxen License²³⁷ (=BSD 3-Clause)</p> <p>MIT License</p> <p>Common Public License 1.0 (CPL 1.0) Creative Commons (CCO 1.0)</p>		<p>producing a derivative work, affected by copyleft²³⁸. The component may not be licensed under Apache, because Apache does not tolerate LGPL software in Apache products²³⁹.</p>	<p>Redistribution requirements (if packaged with dependencies):</p> <p>Codes of dependencies must stay under original licenses. Please keep copyright, disclaimers and license notices intact.</p> <p>EPL, CDDL, CPL, EPL, MPL, LGPL: credit the use of dependencies under these licenses, include the license terms into distribution and inform the user how to get the source code.</p>
14	VPH-HF	HTTP calls, command	HTTP calls:	Command line arguments, HTTP	No binding license.

²⁴⁰ <https://www.gnu.org/licenses/license-list.en.html>

²³⁶ <http://asm.ow2.org/license.html>

²³⁷ <http://jaxen.org/license.html>

²³⁸ “This allows, for example, programs using MPL-licensed code to be statically linked to and distributed as part of a larger proprietary piece of software, which would not generally be possible under the terms of stronger copyleft licenses”, See: <https://www.mozilla.org/en-US/MPL/2.0/FAQ/>.

²³⁹ <http://www.apache.org/legal/resolved.html>

	CINECA/USFD	<p>line calls, procedure calls</p> <p>-----</p> <p>BSD 3-Clause License:</p> <p>-Django²⁴¹: procedure calls</p> <p>-Celery²⁴²: procedure calls</p> <p>BSD 2-Clause License:</p> <p>-Django REST²⁴³: procedure calls</p> <p>LGPL 2.1</p> <p>-Taverna²⁴⁴: HTTP calls</p> <p>LGPL 3</p> <p>-MUSCLE²⁴⁵: command line calls</p> <p>MIT License:</p> <p>-Bootstrap²⁴⁶: HTML, CSS</p> <p>-jQuery²⁴⁷</p>	<p>-----</p> <p>CRAF</p> <p>Model Repository</p> <p>In Silico Trial Repository</p> <p>Clinical Data Repository</p>	<p>calls, procedure calls are communication mechanisms, normally used between separate programs²⁵⁰. Therefore, use of these dependencies by VPH-HF via these mechanisms of communication shall not create any legal implications on the licensing of VPH-HF.</p>	<p>As intended, the component may go open source under Apache v2.</p>
--	-------------	---	--	---	--

²⁴¹ <https://github.com/django/django/blob/master/LICENSE>

²⁴² <https://github.com/celery/celery/blob/master/LICENSE>

²⁴³ <http://www.django-rest-framework.org/#license>

²⁴⁴ <https://github.com/taverna/taverna-server/blob/2.5/LICENSE>

²⁴⁵ <https://www.gnu.org/licenses/lgpl.html>

²⁴⁶ <https://github.com/twbs/bootstrap/blob/master/LICENSE>

		<p>GPL v2²⁴⁸:</p> <ul style="list-style-type: none"> -MySQL: MySQL calls Mozilla Public License 1.1 (MPL 1.1): - RabbitMQ²⁴⁹: AMQP calls 			
15	<p>Vasculature Nephroblastoma</p> <p>Software implementing the vasculature hypomodel for nephroblastoma UOXF</p>	<p>API calls and dynamic linking in all cases. Executable built using Chaste source code</p> <p>-----</p> <p>BSD-3-Clause License:</p> <ul style="list-style-type: none"> -Chaste 251 -VTK252 <p>BSD 2-Clause:</p> <ul style="list-style-type: none"> -PETSc253 <p>Boost Software License:</p> <ul style="list-style-type: none"> -Boost254 	<p>MUSCLE API</p> <p>-----</p> <p>the ICCS Tumour Growth model and FORTH Metabolic model</p>	<p>Dependencies licenses are compatible with each other²⁵⁶.</p> <p>Building executable using Chaste source code does not produce licensing implications, because Chaste BSD 3-Clause License allows modifying the BSD source code and license a derivative under any terms, provided the BSD terms for the Chaste code are met.</p> <p>Interaction with the CHIC components/models via MUSCLE API does not produce licensing implications, because MUSCLE LGPLv3 license, Section 4 allows “<i>combining or linking an Application with the Library</i>” and to “<i>convey a Combined Work under terms of your</i></p>	<p>No binding license.</p> <p>If packaged with LGPL MUSCLE, licensing under Apache might create incompatibility issues with LGPL, because Apache does not tolerate LGPL in Apache products²⁵⁸. Suggested open source licenses: copyleft GPLv3, weak copyleft LGPLv3, lax permissive license: BSD 3-Clause.</p> <p>Redistribution requirements (if packaged with dependencies):</p> <p>Codes of dependencies must stay under original licenses. Please keep copyright, disclaimers and license notices intact.</p> <p>LGPL v3: credit LGPL library, provide LGPL license for the library, inform how to get the source for the library. Please see para 4 LGPL v3 for more details²⁵⁹.</p>

²⁴⁷ <https://github.com/jquery/jquery/blob/master/LICENSE.txt>

²⁵⁰ <https://www.gnu.org/licenses/gpl-faq#MereAggregation>

²⁴⁸ <http://www.mysql.it/about/legal/licensing/oem/#5>

²⁴⁹ <https://www.rabbitmq.com/mpl.html>

²⁵¹ <https://opensource.org/licenses/>

²⁵² <http://www.vtk.org/licensing/>

²⁵³ <https://www.mcs.anl.gov/petsc/documentation/license.html>

²⁵⁴ <http://www.boost.org/users/license.html>

		LGPL v3: Muscle ²⁵⁵		<i>choice, that, taken together, effectively do not restrict modification of the portions of the Library contained in the Combined Work and reverse engineering for debugging such modifications</i> ²⁵⁷	For Chaste/BSD: please keep all license, copyright and disclaimer notices in Chaste source files intact, include the BSD license for Chaste into distribution.
16	Vasculature Lung software implementing the vasculature hypomodel for lung cancer UOXF	API calls and dynamic linking in all cases. Executable built using Chaste source code. ----- - BSD-3-Clause License: -Chaste 260 -VTK261 BSD 2-Clause: -PETSc262 Boost Software License: -Boost263 LGPL v3:	MUSCLE API ----- the ICCS Tumour Growth model and FORTH Metabolic model	Dependencies licenses are compatible with each other ²⁶⁵ . Interaction with the CHIC components/models via MUSCLE API does not produce licensing implications, because MUSCLE LGPLv3 license, Section 4 allows “combining or linking an Application with the Library” and to “convey a Combined Work under terms of your choice, that, taken together, effectively do not restrict modification of the portions of the Library contained in the Combined Work and reverse engineering for debugging such modifications” ²⁶⁶	No binding license. If packaged with LGPL MUSCLE, licensing under Apache might create incompatibility issues with LGPL, because Apache does not tolerate LGPL in Apache products ²⁶⁷ . Suggested open source licenses: copyleft GPLv3, weak copyleft LGPLv3, lax permissive license: BSD 3-Clause. Redistribution requirements (if packaged with dependencies): Codes of dependencies must stay under original licenses. Please keep copyright, disclaimers and license notices intact. LGPL v3: credit LGPL library, provide LGPL license for the library, inform how to get the source for the library. Please see para 4 LGPL v3

²⁵⁶ <https://www.gnu.org/licenses/license-list.en.html>
²⁵⁸ <http://www.apache.org/legal/resolved.html>
²⁵⁹ <https://opensource.org/licenses/LGPL-3.0>
²⁵⁵ <https://www.gnu.org/licenses/lgpl.html>
²⁵⁷ <https://opensource.org/licenses/LGPL-3.0>
²⁶⁰ <https://opensource.org/licenses/>
²⁶¹ <http://www.vtk.org/licensing/>
²⁶² <https://www.mcs.anl.gov/petsc/documentation/license.html>
²⁶³ <http://www.boost.org/users/license.html>

		Muscle ²⁶⁴			for more details ²⁶⁸ .
17	Universal growth response treatment UNITO and to	From scratch	UPENN Nephroblastoma Molecular Model	N/A	No binding license. Suggested open source license: Apache v2

²⁶⁵ <https://www.gnu.org/licenses/license-list.en.html>

²⁶⁶ <https://opensource.org/licenses/LGPL-3.0>

²⁶⁷ <http://www.apache.org/legal/resolved.html>

²⁶⁴ <https://www.gnu.org/licenses/lgpl.html>

²⁶⁸ <https://opensource.org/licenses/LGPL-3.0>

18	Molecular Network Model UPENN	<p>Calling object code</p> <p>-----</p> <p>COPASI: artistic license 2.0²⁶⁹</p>	<p>Command line</p> <p>-----</p> <p>the CHIC data repository (CDR) to take the minML files containing the molecular data. This is invoked at the command line by specifying flags.</p>	<p>No license incompatibility issues.</p> <p>Internal interaction with the CHIC components/models does not produce licensing implications, because communication via command line arguments is a method of communication usually used between two separate programs²⁷⁰. When components interact in this way their licenses are independent from each other.</p>	<p>No binding license²⁷¹.</p> <p>Distribution in object code (on commercial basis) and open source possible.</p> <p>Suggested open source license: Apache v2</p> <p>If packaged with COPASI, COPASI must stay under artistic license and distribution requirements of artistic license must be observed.</p> <p>Redistribution requirements (if packaged with COPASI):</p> <p>Artistic license, 2): “you duplicate all of the original copyright notices and associated disclaimers. At your discretion, such verbatim copies may or may not include a Compiled form of the Package.</p> <p>(3) You may apply any bug fixes, portability changes, and other modifications made available from the Copyright Holder. The resulting Package will still be considered the Standard Version, and as such will be subject to the Original License.”²⁷²</p>
19	Biomechanical	<p>Calling executables</p> <p>-----</p>	<p>Command line arguments, MUSCLE</p>	<p>GPL considers: “Linking a GPL covered work statically or</p>	<p>Component license: GPLv3 (if</p>

²⁶⁹ <http://copasi.org/Download/License/>

²⁷⁰ <https://www.gnu.org/licenses/gpl-faq#MereAggregation>

²⁷¹ Artistic license 2.0:“(7) You may aggregate the Package (either the Standard Version or Modified Version) with other packages and Distribute the resulting aggregation provided that you do not charge a licensing fee for the Package. Distributor Fees are permitted, and licensing fees for other components in the aggregation are permitted. The terms of this license apply to the use and Distribution of the Standard or Modified Versions as included in the aggregation. (8) You are permitted to link Modified and Standard Versions with other works, to embed the Package in a larger work of your own, or to build stand-alone binary or bytecode versions of applications that include the Package, and Distribute the result without restriction, provided the result does not expose a direct interface to the Package.” See: <https://opensource.org/licenses/Artistic-2.0>.

²⁷² <https://opensource.org/licenses/Artistic-2.0>.

simulator UBERN	<p>Apache v2: - Apache Xerces XML Parser²⁷³ GPL v2: - CodeSynthesis XSD²⁷⁴ BSD 3-Clause License: - VTK²⁷⁵ GPLv3/LGPLv3 - CGAL²⁷⁶ LGPL v3: - MUSCLE 2²⁷⁷ MIT License: - easylogging++²⁷⁸ MIT License (MIT) In addition, BMS calls an external programme (executable): - FEBIO license²⁷⁹</p>	<p>API</p> <p>-----</p> <p>input:</p> <p>-command line arguments (numeric parameters and paths to files, such as configuration file and image files)</p> <p>-MUSCLE API (input from OncoSimulator)</p> <p>output:</p> <p>- MUSCLE API (input for OncoSimulator)</p> <p>- 3D image/geometric model files (not used in</p>	<p><i>dynamically with other modules is making a combined work based on the GPL covered work. Thus, the terms and conditions of the GNU General Public License cover the whole combination</i>²⁸⁰.</p> <p>Apache v2, LGPLv3, BSD 3-Clause, MIT License are GPL-compatible²⁸¹.</p> <p>If packaged/distributed with GPL dependencies, the component must go under GPLv3.</p> <p>Communication via command line arguments and MUSCLE API does not produce licensing implications.</p>	<p>packaged/distributed with GPL dependencies)</p> <p>Distribution in source code and object code allowed. Distribution in object code must be accompanied with the possibility to get the source. Non-commercial licensing only. Charging royalties not permitted. Fees for physical distribution and warranties/servicing allowed²⁸².</p> <p>If packaged/distributed with dependencies, dependencies must stay under their own licenses. Please keep copyright, license and disclaimer notices in the dependencies files intact. Please label the use of LGPL library and provide instruction how to get LGPL source code. Please see para 4 LGPL v3 for more details²⁸³.</p>
--------------------	--	--	--	--

²⁷³ <https://xerces.apache.org/xerces-c/>

²⁷⁴ <http://www.codesynthesis.com/products/xsd/>

²⁷⁵ <http://www.vtk.org/>

²⁷⁶ <http://www.cgal.org/project.html>

²⁷⁷ <http://apps.man.poznan.pl/trac/muscle>

²⁷⁸ <https://github.com/easylogging/easyloggingpp>

²⁷⁹ University grants to you a non-exclusive, non-transferable right to use the SOFTWARE in a single installation on a single machine in a single geographic location for Non-Commercial Use. Recipient agrees not to use such SOFTWARE for any Commercial Purpose.

Use of the software for any Commercial Purpose means use of Software by you for direct or indirect financial, commercial or strategic gain or advantage. See: <http://febio.org/download/febio-license/>

²⁸⁰ <https://www.gnu.org/licenses/gpl-faq#GPLStaticVsDynamic>

²⁸¹ <https://www.gnu.org/licenses/license-list.en.html>

²⁸² <https://opensource.org/licenses/GPL-3.0>

²⁸³ <https://opensource.org/licenses/LGPL-3.0>

			CHIC)		
20	Preprocessing Tool FORTH	Calling object code: ----- Apache v2: - SimpleITK ²⁸⁴ BSD 3-clause license: - NumPy ²⁸⁵ PSF license: - Python ²⁸⁶	Command arguments ----- Any hypermodel Oncosimulator	Licenses of dependencies are compatible with each other ²⁸⁷ . Internal interaction with the CHIC components/models does not produce licensing implications, because communication via command line arguments is a method of communication usually used between two separate programs ²⁸⁸ . When components interact in this way their licenses are independent from each other.	No binding license. Suggested open source license: Apache v2 Distribution in object code on proprietary basis and open source allowed. If packaged/distributed with dependencies, the dependencies must stay under original licenses. Please keep copyright, disclaimers and license notices intact.
21	Metabolic Model FORTH	Calling object code: ----- LGPL v3: -MUSCLE 2	MUSCLE API ----- Hypermodels: - Nephroblastoma Multimodeller hypermodel -Lung Multimodeller hypermodel Hypomodels: - ICCS Oncosimulator	Linking to LGPLv3 MUSCLE and interaction with the CHIC components/models via MUSCLE API does not produce licensing implications, because LGPLv3 license, Section 4 allows “ <i>combining or linking an Application with the Library</i> ” and to “ <i>convey a Combined Work under terms of your choice, that, taken together, effectively do not restrict modification of the portions of the Library contained in the Combined Work and reverse</i> ”	No binding license. If packaged with LGPL MUSCLE, (a) licensing under Apache might create incompatibility issues, because Apache does not tolerate LGPL in Apache products ²⁹⁰ ; (b) license must allow modification and reverse engineering for the customer’s own use (para 4 LGPLv3). Suggested open source license: weak copyleft LGPLv3, Apache v2 (if packaged without MUSCLE), BSD 3-Clause. Redistribution requirements (if packaged with

²⁸⁴ <http://www.simpleitk.org/SimpleITK/project/license.html>

²⁸⁵ <http://www.numpy.org/license.html>

²⁸⁶ <https://docs.python.org/3/license.html>

²⁸⁷ <http://www.apache.org/legal/resolved.html>

²⁸⁸ <https://www.gnu.org/licenses/gpl-faq#MereAggregation>

			- UOXF Vasculature	<i>engineering for debugging such modifications</i> ²⁸⁹	dependencies): Codes of dependencies must stay under original licenses. Please keep copyright, disclaimers and license notices intact. LGPL v3: credit LGPL library, provide LGPL license for the library, inform how to get the source for the library. Please see para 4 LGPL v3 for more details ²⁹¹ .
22	Hypermodelling Editor FORTH	Calls object code: ----- Eclipse Public License 1.0: - Clojure ²⁹² - clojurescript ²⁹³ - H2 database ²⁹⁴ - ring-server ²⁹⁵ - compojure ²⁹⁶ - hiccup ²⁹⁷	HTTP calls to: ----- - Model Repository - VPH-Execution Framework - Metadata Repository	Licenses of dependencies are compatible with each other. Communication via HTTP calls does not produce licensing implications.	No binding license. Distribution in object code (on commercial basis) and source code possible. Suggested open source license: Apache v2 Redistribution requirements (if packaged with dependencies): Codes of dependencies must stay under original licenses. Please keep copyright, disclaimers and

²⁹⁰ <http://www.apache.org/legal/resolved.html>
²⁸⁹ <https://opensource.org/licenses/LGPL-3.0>
²⁹¹ <https://opensource.org/licenses/LGPL-3.0>
²⁹² <http://clojure.org/index>
²⁹³ <https://github.com/clojure/clojurescript>
²⁹⁴ <http://www.h2database.com/>
²⁹⁵ <https://github.com/weavejester/ring-server>
²⁹⁶ <https://github.com/weavejester/compojure>
²⁹⁷ <https://github.com/weavejester/hiccup>

		²⁹⁸ ²⁹⁹ ³⁰⁰ ³⁰¹ ³⁰² ³⁰³ ³⁰⁴ Apache License Version 2.0: - http-kit ³⁰⁵ MIT: - reagent ³⁰⁶ - re-frame ³⁰⁷ - ring ³⁰⁸ BSD: - prone ³⁰⁹			license notices intact. Provide instructions to the user where/how to get the EPL source codes.
--	--	---	--	--	---

²⁹⁸ <https://github.com/reagent-project/reagent-forms>
²⁹⁹ <https://github.com/reagent-project/reagent-utils>
³⁰⁰ <https://github.com/JulianBirch/cljs-ajax>
³⁰¹ <https://github.com/clojure/data.json>
³⁰² <https://github.com/clojure/java.jdbc>
³⁰³ <https://github.com/gf3/secretary>
³⁰⁴ <https://github.com/weavejester/environ>
³⁰⁵ <https://github.com/http-kit/http-kit>
³⁰⁶ <https://github.com/reagent-project/reagent>
³⁰⁷ <https://github.com/Day8/re-frame>
³⁰⁸ <https://github.com/ring-clojure/ring>
³⁰⁹ <https://github.com/magnars/prone>

23	CHIC platform	CHIC components	http calls, command line arguments, procedure calls, exchange of data files, etc.	<p><i>“pipes, sockets and command-line arguments are communication mechanisms normally used between two separate programs. So when they are used for communication, the modules normally are separate programs....An “aggregate” consists of a number of separate programs, distributed together on the same CD-ROM or other media.”.</i>³¹⁰</p>	<p>Aggregate</p> <p>The platform license should not limit the license terms and rights provided by individual licenses of the components.</p> <p><i>The GPL permits you to create and distribute an aggregate, even when the licenses of the other software are non-free or GPL-incompatible. The only condition is that you cannot release the aggregate under a license that prohibits users from exercising rights that each program's individual license would grant them.</i>³¹¹</p>
----	---------------	-----------------	---	---	---

³¹⁰ <https://www.gnu.org/licenses/gpl-faq#MereAggregation>

³¹¹ <https://www.gnu.org/licenses/gpl-faq#MereAggregation>