Title
Novel informatic platform for the standard unification of donors, offspring and patients records: institutional registry requirements in the reproductive biomedicine regulations compliance.

Mauro Costa a,*,1, Salvo Reina b, Mauro Giacomini c

a) S.S.D.Medicina della Riproduzione, Ospedale Evangelico Internazionale, Genoa, Italy
b) DSCDI, DISC Faculty, School of integrated Medicine, University of Genoa
c) DIBRIS – Department of Informatics Bioengineering Robotics and System Engineering, CEBR – Center of Excellence for Biomedical Research, University of Genoa

* Corresponding author. Email address: mauro.costa@oeige.org
1 Present address: S.S.D.Medicina della Riproduzione, Ospedale Evangelico Internazionale; Corso Solferino, 1A - 16122 Genoa, Italy

Abstract
The Donor ART (DART), recently promulgated by Italian institution, has posed several incumbent questions to the Artificial Reproductive Techniques (ART) world. Besides the critical social and ethical aspects, a crucial issue is the engagement of a centralised “Donors-offspring Registry” which introduces new stringent criteria up on the treatments of the sensitive data of donors, newborns and patients. Such registry will have to comply with applicable Italian laws (Legge 40/04, Dlg 191/07 and Dlg 16/10, the so called “Patto per la Salute 2014-2016) and the European Tissue Directive (2006/17/CE, Ed. 2004). Moreover, recent Guide Lines were introduced by the Italian Conferenza Stato Regioni in the 4/9/2014 which officially refer to the creation of dedicated “donors-offspring” registry for the fecundation of donated gametes issuance. Lastly, the entire model will have to comply with a preliminary check of the Privacy Authority (Dlgs196/03). In order to fulfill the mandatory requisites related to anonymisation, traceability, protection of biomedical data and creating an informatic “platform design” capable of an heuristical objective discrimination and control of donors’s phenotypes, we coupled cryptology and Artificial Intelligence (AI) techniques to design and propose an adequate software architecture of such launching registry. At the time of our study, no official standard was unified to this purpose nor scientific works oriented on DART Registry specification are available yet, therefore, we worked on a novel central IT repository conception capable of accomplishing all the essential compliances with law and regulation. Our aim is to describe technical and functional features of an informatic data-banking model for the DART and design a working prototype which is indispensable to its realisation. We also candidature an entity-relational structure of the database at a systemic level of abstraction under the upcoming technological paradigm of the Cloud Computing and Big Data technologies. This vision was demonstrated to be particularly suitable for the ART with gamete donation because of its peculiar features such as security multi-layers control, segregation of accountability, central algorithm for control logics and others. More in general, we theorised a biomedical semantic informational storage implementation as a possible reference model for the unification under an institutional nationwide mandatory standard for the DART. This paper is intended for a divulgative description of the forthcoming software technology scenario which will impat on the reproductive biomedice and it can not be dodged: in the close future embryologists and physicians belonging to the ART professionals community should become acquainted with insight and notions pertaining informatic tools and computation logics as a daily practise of their ART expertise. We also provide system IT administrators and programmers who works for ART Centers, with essential technical informatic hints of the software design and infrastructure of the platform model.

Keywords: Artificial Reproductive Techniques (ART), Donor gamete, Donor-Offspring Registry, Clinical Data Protection, medical phenotypes discrimination, computer identity control, quality and security assurance, Biomedical Privacy Compliance
Introduction

The ART is an articulated biomedical process which is daily conducted by ART centers institutionally accredited as Tissues Institutes which treat gametes with procreative purposes on sub-fertile patients population. The recent Italian re-introduction of the hetelologous gamete donation, better known as donor ART (DART in this paper), does not substantially changes the biomedical process while, on the contrary, it strongly impacts on the way normative and regulative prescriptions of safety and security enforce tight controls on the donor’s identity, his/her health status and biological integrity of reproductive cells. Since the 2003, our group has electronically registered ART’s cycles records according to the applicable legislation and norms to guarantee the uttermost level of standard concerning privacy, data protection and management of the emergencies due to possible adverse events. Because of the upcoming DART changes, we felt the need to revise our informatic system by reconsidering compliance and additional requisites needed to leverage the internal standard. We took this works as an occasion to investigate a more widely adoptable software model in view of the official registry for donors and offspring institutionalised by law.

The Table 1 summarises the major new topics implied by DART and each of these topics it represented to us a Critical Control Point (CCP) with a brief description of the related drawbacks and hichs which we decided to be focused on. For sake of completeness, it is needed to be said that together with the ones that we have considered, other critical aspects concerning social, legal and jurisprudential implication were also defined by the Italian Institutions. For instance, and just to mention two relevant questions, economic costs for the patients and annexing of DART procedure among the socially acknowledged services of the Italian Sistema Sanitario Nazionale (SSN) according essential levels of assistance defined LEA (Livelli Essenziali di Assistenza) by institution.

Ultimately, our interest it was concentrated on the donor-offspring registry and its related topics to prepare and anticipate ART centers in adequating their data treatment. Most of the enquiries that we undertook to solve each of the CCP’s controversy, consisted of a simple updated appliance of our currently implemented databank software, while in a minor number of cases we opt for an entire re-engineering (i.e.: technical cascade code design) of the software. Finally, a few yet crucial features, required an highly specialized and suffisticated software’s skill, hence we involved the engineers of the DIBRIS of the University Genoa: a group already specialised in medical informatics. This interdisciplinary approach it was essential to confer to our model the possibility to overcome stringent assumption of the Italian Privacy Authority denial towards the Central Registry permissions.

The Table 1 is by fact the sequential check-list of our work: each of the challenge issue was studied and subsequentially integrated with the other critical points until an accettable solution was found for all of them. Finally, the plethora of solutions was harmonised into a coherent framework model. We hereby share with colleagues our efforts in studying each trouble associated one-by-one with the CCPs and report the possible IT solution to achieve data preservation/treatments and perform quality and security adequacy. Because this communication is primarily addressed to ART professionals and colleagues, we wrote separate paragraphs and reported our findings and proposals as a conceptual and iconographical description: we thus privilege the clinical and biomedical audience to improve the contents clarity. The necessity of an heated debate between biomedicine and informatics bearad us to revise the ART procedure on a diverse prospective, a stimulating opportunity of improvement.

Materials and Methods

ART data repository

The first step of our study consisted in achieving a complete database with a consistent and realist number of ART records. Since the 2003 the group of our Reproductive Medicine Center of the OEIGE hospital in Genoa has systematically registered clinical and biological data of the patients’ ART cycles. The former proprietry software was developed by authors and based on SCRUM agile framework (7,8) and Object Oriented Modelling and Agile Programming OOM-OOP standards. The application inner logics implemented more than 300 heuristical rules by developed by authors and based on SCRUM agile framework (7,8) and Object Oriented Modelling and Agile Programming OOM-OOP standards. The application inner logics implemented more than 300 heuristical rules by
Software technology and standards
The model we propose is technically classifiable as a platform, therefore the system software, globally considered, is a “chimera” of several specifications in the Information Communication Technologies (ICT aka IT). We have used either proprietary or open source platforms for the implementation of working-logic and the testing of the proof-of-concept interface. As far as for the data-base architectural design is concerned, we adopted an open-standard approach so that in the future the theoretic modelling could be easy applicable and formalised with any kind of database product. At last, details of the algorithms’s models already developed by the authors, get over the scope of this context and can be found elsewhere (1,2,3).

Authentic and genuine data repertoire
It was essential to us the handiness of an authentic collection of ART clinical cases. Thank to this relational database we were able to specifically address some of the problems in the anamnestic and personal data profiles. More specifically, we were able to characterise a pattern of heuristic rules which allowed to compare and compute a similarity index (taximetric phenetic distance) between cases and, within the same couple for diagnostic and embryological data follow-ups. A preliminary statistical trial was performed on the repository to determine the canonical power test and the non-parametric multi-factorial variables distribution. The PCO/PCA (Principal Components analysis) statistics used for cluster analysis were carried out with PAST and JMP software product (see C3, C4). Physicians and programmers hence identified the proper fields panel to distinguish clusters on which to apply discriminant analysis and UPGMA (Unweighted Pair Grouping Matching Analysys).

Prior than using the “raw-data” records (unbiased and original variable values) for the evolved and more complex task of the AI techniques, personal references such as names and surnames, where “fantasised”, meaning by this that original birth and family names where substituted with common epithets derived from cartoon and books characters with a large reputation (i.e. : Harry Potter, Fata Morgana, Mister Magooo ect.). After that database was exported outside the hospital to be processed by the informatic laboratory.

DART Knowledge Base propaedeutic notions for heretics
It is diffused the concept of the database as a collection of tables and records (lines of a table). Each record has fields (or columns) which contain data (dates, descriptive strings and numerical values). The tables are then linked with a so called “join condition” (technically “association key”). The ensemble of the tables together with all the meta-information of the correlation keys is generally known as logical schema of the database (usually derived from the conceptual model agreed with the medical group with the Entity-Relational diagram). This generic scheme is implemented within a commercially available Relational Database Management System (RDMS) which allows the interaction with data stored via a standardized and Structured Query Language (SQL). These queries are able to select and extract specific record’s subsets. In our model this scheme is further advanced by the addition of a so called Heuristic Layer which is an additional software component of the system responsible for a set of “rules” and meta-variables which monitor and proactively control the coherence and the integrity of data across tables, and eventually multiple databases.

This second scheme is more properly called “Knowledge Base” and is a more adequate model for the purpose of our studies for the DART-KB. Later on in this paper, we shall give on-practice examples and reuse latter concepts in the appropriate context for a further logical step in an evolved software technology.

The Table 3 lists an overall outlook of the ER database that we have used: its description is intentionally reported at a logical and generic level because in this context we wish to emphasise the typed segregation and work-flow of the architectural structure. Such a structure, shows how the repository design follows and correlates with typical ART methodology tasks. A complete list of the fields is not necessary here, yet a more detailed analysis of specific fields and variables can be appreciated later on in this paper once we shall get to the specification of the Heuristic Engine Controller (HEC) responsible for the control logic implementation of the Pattern Matching recognition model (2). In the Figure 3 the cascade charting aims to simplify the interpretation of the association of the ER structure of the Table 3 with the sequence of informational data-entry and storage evolution.

The identity code dilemma
Since early times in electronic databanks, especially those concerning Health Care and medical records, a crucial feature of the software is the robustness which assure the record identification mechanism. Because of the italian decree Dlg191/07, in the ART the “identification” pertains gamete sample, diagnostic result, material and instrument classification, cultured cells, reagents and, most notablyworthy, the person identity. We here delineate this latter aspect in that is critical either for confidentiality, protection, and ethical influence; yet, when we found the computational solution we adopted pragmatically for all other Identity Classification purposes.

When we refer to a patient, a donor and a newborn, it is essential to uniquely identify the physical person beyond any possible doubt. On the informatics, the classical approach of Database Administrators (DBA) for Public Health registry has used the italian “Codice Fiscale” (CF), a sort of the SSN (Social Security Number) in the USA. Unfortunately, the CF code is inherently weak because does not guarantee uniqueness. Practically speaking, if taken alone, any information derived from official documents, can not satisfy all ART expectation.
The identity code is pivotal to all the electronic data processes because, once is defined, it will be implemented for all the traceability routines. Moreover, in some cases the code generated with “hash” techniques (see later on for details), needs to be associated with a “time-stamp seed” (the italian “marca temporale”). For instance, this is the case of semen, embryos and the oocytes samples in the Cryopreservation environment as well as for the Access Control List (ACL) enforcement for the operators that can gain control on the ART database. Virtually any CCPs and phases shown in Table 1 and in Figure 4 can receive benefit from a unified hash coding algorithm. A better qualification of our hashing methods will be addressed more specifically in the Result section below.

**Admission and enrolling heuristics**

It is worth to remind that because of the premise claimed in the introduction, we have used the concept of Identity Disclosure not pending on the potential benefits or concerns for DART subjects to be reciprocally informed (recipient, donors and offspring). This pertains ethics consideration on the egg and sperm donation matter while we concentrated our efforts in the neat separation of personal data. All the heuristic rules and the automated controls of the DART-KB model are aimed to the avoidance of any future contacts between DART “actors”. In the registry we modelled, this philosophy starts from the very beginning by validating information at the data-entry levels. In short, since the early stages of the demographic and pre-conceptional mandatory serological test results, the two patterns are checked for their completeness and suitability. If confirmed, the cascade of the HCE allows the remote physician who is currently inserting information into the DART Registry to proceed to the next step (duties accountability validation mandatory by law).

The Figure 4 shows the inferential flow-chart which schematically explains the status control logics of the HCE mentioned above. In the DART-KB model, this mechanism accounts for the constant association and audit validation control which ensure the linkage between a CCP and its Access Control List (ACL) implying by this the correct cascade (south oriented arrow) of the phases which eventually intercept any anomaly and cease further data processing.

**Data security & privacy measures**

In the “real world” of DART, the concept of Data Protection concurrently means confidentiality, availability, integrity pro-active measures. All these aspects need to be guaranteed to preserve and secure patients and donors information. A recent pronouncement of the italian Authority for the Protection of Personal Data (“Comunicato Stampa Garante Privacy” Sept 10th 2014) specifically addressed new stringent campaign denominated “Privacy Sweep 2014” for the control and the enforcement of privacy standard in the Health Care and in the medical software apps on internet called Green paper on Mobile Health by Global Privacy Enforcement Network, GPEN (9). The measures dictated on the guidance impact on logical, physical and management activities of the operator in the ART centers, hence the hypotesis of a central registry needs to be carefully considered at the design time of the project (Privacy by default and by design according to European Parliament and Council data protection reform 95/46/EC, 2009).

In our model this is accomplished by systematically implement cyphering with cryptology at the fields level of the DART-KB. This means that even under a dramatic adverse situation of database’s theft, the contents will not be intelligible outside the native environment.

Technically, chipher techniques use transformation of a value (scrambling, masking, recoding) globally defined “hashing”: such a method uses an encoding algorithm associated to a “seed” (encryption key). Hashing techniques are largely used in biomedicine (10).

In the model of KB-DART we added an additional care: in fact, the seed of a single variable in the record uses the Unique Identifier (UI) of the table it belongs to (unique key field) and in turn, each UI will be associated to the ART Center which originated the record at the time of first data-entry.

Since it is possible to achieve a complete view of a data profile only by accessing simultaneously segregated tables, it is practically impossible to decipher original values. Each table has different seeds. This technical modelling specification contributes to satisfy the CCPs in the Table 1

**Biometrics anti-fake strategy**

In the DART-KB we considere of particular interest the presence of BLOB (Binary Large Object) fields for each of the informational object associate with anamnestic and personal identification. This is because karyotype and TAC-PCR genetic tests are very expensive and its availability is not ubiquitous at all, while small biometrical devises (fingerprints, retinal scanners etct.) can be easily interfaced with a database system.

On practise, small and cheap biometrical instruments are already available and usually are provided with USB and/or wireless connection to a computer (wifi and bluetooth). The layout can be stored inside a record either as image or normalised string hashing (11), thus it will be possible to compare and confirm a person identity even in the case of a false or incomplete personal document information. Interestingly, inside the file structure of an image (matrix transposition of vectorial or raster bitmap), we also took advantage by the so called “steganography method” to store additional meta-information data (12, 13). In this manner we mitigated if not eradicated the risk exposure to identity theft and person substitution.

As previously stressed out, one of the major concern of the DART is the necessity of being able to prevent a donor from using multiple samples in multiple centers. To accomplish this law’s requirement the Registry will have to provide a
trustworthy way of control which can recognise a donor beyond his ID card document with the better as possibile certainty. In a very much the same way, institutional guide lines assume that ART operators (either technicians, biologist or physician) will not be donors and will not interfere with any donor record's information if cross-related as relative to patients. Even more clear is the counter-indication for a physician within the same ART center.

In order to resolve these latter situations, the basic feature across the entire levels of the HEC is a “strong authentication”. In the Figure 6, it can be appreciated that the yellow box with the HEC mediated access to patient’s records it is bound through the “intramural sourcers” to the virtuos cycle of accountability and governance namely “Groups Hypervision” and “Security & Privacy” grey ellipses.

The strong authentication is not only related to the Identity Code already explained above, yet it also involves the protocols matching communication of the single transaction. In the DART model we believe this can be accomplished through Asymmetrical double Public-Private Certificate key Cryptography. This standard (13) is to us particularely suitable to the authentication requisites of the ART work-flow because the protocol is designed for distributed Trusted Platform Module (TPM) and is also implemented to support token-interface of Hardware Security Module (HSM). This solution is desirable mainly for two reasons : 1) all the critical data-entry and data-editing procedures of the operators on the ART records can be logged and controlled and, 2) at a connection security level, the postulated Cloud Computing network already provides the so called API capabilities (Application Programming Interface) for the use of “tokens”. A token is a personal device used to be authenticated and authorised to log-on in a system very much like the case of the “key-generator” device that we use to get into a money bank account on internet.

Cloud Computing revolution for DART-KB
All the components of the DART-KB platform described so far were conceived as modular parts of a wide model and each module was designed and structured to tackle and find a solution for a specific or a group of requiring of the CCPs in the DART constraint. As already written previously, when globally considered, our model is a software platform because it contains several type of software elements (protocols, algorithms, application, libraries, service, processes and database). Metaphorically, all the elements concur in formalising the technical specification of an articulated system maden by objects, bolts and gears of a wider machinery. The entire “apparatus”, in turn, needs to be conjectured in a “real world” scenario, namely the future realisation of an institutional registry, therefore we also studied a plausible and practicle central infrastructure to host the ICT DART system.

The present study implied a continuous “knowledge osmosis” between medical requests of physicians and informatic staff responsible for the software implementation. Once we got to the point in which were reciprocally identified and understood, the concepts related to the Database and the Knowledge-Base migration, we faced a second and more vast step of the Cloud Computing (CC). The dialogue between biomedical staff and highly educated informatic skills, finally got a clou by simplifying the picture and explaining the CC as it was a “big container” of the entire DART-KB system. The CC indeed, is a revolutionary paradigm in the Informatic and Communication Technology (ICT) and is changing the future of the data computation in any field at rapid pace. This phenomenon is investing all industrial, production and academies at a planetary level: its complete treatise is not even approachable in this context, thus we describe in an essential way how the model for the DART-KB should be integrated and harmonised with CC data center.

The CC uses internet to connect a wide number of remote computers to a central virtual Data Center. This is exactly what would be desirable for an institutional Registry for DART. If compared with the conventional Data Center the a CC network offers several advantages, especially in the biomedical and Public Health Information Technology: data security (segregation of duties, traceability, data protection and automated backups preservation among the others) and shared communication knowledge (i.e.: distributed databank, university networks, E-learning systems) are the most noteworthy features.

Essentially, the CC can be Public, Private or Hybrid and has three major architectural models : IasS, PasS and SaaS (Infrastructure, Platform and Software as a Service respectively). In the case of our platform model for DART we think that a Private CC is the preferable choice. Obviously, the term Private does not refer to the commercial acceptation, it rather means that the governance of the Central network is controlled by the subject who owns the data center institutional premises where physically the Cloud infrastructure resides.

An important notion concerning the applicability of the CC in the DART-KB finalised for the institutional Registry, is the “CC hypervisor” (CCHV). On a conceptual point of view, the CCHV is the central control core entirely based and distributed on a Cloud network through internet. More practically and visually, the CCHV is a dashboard (a screen console) accessible by technical and medical administrators which regulate and monitor on real-time the functionality of registry. In our case the system allows ART centers to be connected with the institutional repository with a web interface according to a WEB 2.0 social metaphor. All the features needed for data storage, treatments and control logics are provided by a DART Community Server.
Results

Records Phenotypes and "entity" identification

Because of the theoretic approach of the modelling of the registry, the tangible results can be evaluated in terms of similarity index achieved by the comparison of the Code-strings obtained by the transformation algorithm previously described. The ability of the pattern matching to automatically identify same identities (DART records) is the base functionality scope of the discrimination of two ART donor records.

What is important for a full understanding of the model is the concept of "entity". In the software technology postulated by our system, identification coding techniques are used for any element of the platform; such methods can refer invariably to a gamete sample characteristic, to a phenotype, to an ART operator who is using the DART-KB as well to a pattern of serological / genetical markers.

The most compelling result achievable for a medical records ID-management platform such the one described so far, has to deal, solve and implement all the requisites and drawback specific for the real-world application not forgetting the numerous legal binds of the donors law’s stringency. All the theoretic models have to be concretely applied and implemented. To this scope, the choice of the IT technology of the Content Management Database System (CMDBS) results out from a sophisticated balance of features of the software freely available from scientific community. The prominent characteristic of the K-DART is the entirely new design of the DataBase engine interface and programmability.

Based on what is being stressed out above, ordinary Database Server can not appropriately guarantee the implementation and the compliance required by Donor ART problems. The classical Tables-Record-Field clomunary approach can not fullfill the dynamic and multi-entity ID-Anonimity in the way we assumed in the Entity-Relation scheme conceptually depicted in Tables 3. We have used a table/column approach to simplify the explanation of the functional classification of several variables concerning donors, yet for the realisation of an on-the-field Registry, the abstraction layer of a standard DBMS needs to be outclassed. In essence, orthodox DB engine will not ever work for Donor ART.

In such a context, we oriented the choice towards an innovative type of standard Distributed Content Management System (DCMS) generally denominated either NoSQL or GraphDBs or ObjectDB. The benefit of this type of IT technology are multiple. First of all, many academical institution and community of the Open World render available SDK, support and code libraries of their technology. Secondly, out standing performance can be achieved in querying large datasets if compared with conventional DB Servers.

The third and most crucial feature is the flexibility with which each single information unit (atomic variable) can be put on relation with any other. This means that separated logical procedures and/or cryptology’s computation can be implemented at the most complex and inner level of the data validation. Even more exciting, this new generation of software can be integrated with our own code thanks to the API released by the authors for any thinkable type of language and IT environment.

The Hadoop standard such as Neo4J, Cassandra, MongoDB (just to mention a few), are all suitable and impressive: it is not a coincidence indeed, if this are the current standards in the world of the BIG DATA. A description of these vast domains is clearly out of the competence of this article, therefore we suggest essential cross-reference citations for those who want to deeply investigate (15).

The overall model picture

After the systematic analysis of the critical aspects of the DART biomedical data system solution we are able to summarise a more wide outlook of the entire model. The Figure 6 schematically draws the technological elements and the articulated hierarchy of relationships to which processes and praxis concur and correlate to ART modules of demographic, biomedical, diagnostic and scientific data concerning patients and donors.

One peculiar aspect of the general landscape appreciable in the Figure 6 is the segregation level of the extra/intranet layers with which the model guarantee the control and the compliance with the privacy request of the Authority in terms of ability to block and alert any operation potentially detrimental. Whether an outlier diagnostic assay, an ambiguous identification of a donor identity, an adverse incident concerning a sample manipulation, or simply multi-attempts of informatic access to the system occur, the alert HEC immediately triggers secret communication to GRC staff and it harms the system to prevent violations or incident escalation.
Discussion
The present study provides a model for functional and architectural software platform for a Donor ART Registry. Besides theoretic informatic methodology, the systematic approach to the regulation requirements led us to formalise a ste-by-step solution which could be exported and received by ART professionals.

Central DART Registry relevance in the “medical act” controversy
Simply because it can not be assimilated to the natural human reproduction, the DART have to be considered as a “medical act”. In the first case we refer to a “couple” while in the second, more properly, we use “donor” and “recipient” definitions. The “medical act” assertion itself, implies several responsibilities arising in the intermediary process, meaning by this that either persons, procedures or instrumental device can interfere and modify the course of natural events. For instance, the couple mentioned few lines above, can consciously accept the risk that the offspring will have malformation. Such an event would be considered as a fateful chance of the destiny. On the contrary, in ART the simple reproductive cells manipulation of biologist it could be indicted as potentially concurrent cause.

In the DART, the genetic counselling plays a relevant role because in Italy 1 out of 25 individual has a thalassemia trait and 2% of the population is the epidemiological incidence of Cystic Fibrosis (CF). It is quite obvious that a medical advice must carefully evaluate the donor. Also, if during the periodical control a donor turns to be serological positive for HIV, an immediate process have to be triggered in order to block all the samples cryopreserved and alerting all the ART Centers which might have already used it. These undesirable possibility increase notewortly the need of a DART Registry based on the heuristal network like the one we describe.

Italian Privacy Authority compliance
A special emphasis was reserved to the mechanism with which we assumed the DART Registry should implement centralised heuristical rules in order to comply with all the privacy requests of the italian Authority. In the decree Dlg.196/03 as well as in the future EU reform for the Data Protection, specific articles and clauses are demanding technological, logical and managements praxis for the multicenters medical database and systems. Because in Italy the Privacy Authority disciplines the approval of a new Registry, we stressed the specification of the GRC (Governance, Risk and Compliance analysis) from an ICT point of view by defining a rigorous cascade rules for the enforcement of the “sensitive” data treatments by ART Centers (9).

Standard dissemination
In order to facilitate the major divulgative appliance, the architectural design of the model was based on an entity-relational database structure. This assumption, greatly augments the possibility for the practicle realisation and the technical feasibility of a Registry because the architectural design could be formalised in matter of months; similar consideration would be confirmed for a deployement planning phase.

Hopefully, our experience offer a premising pattern of features inheritable by a nationwide DART registry model so that it could be possible to realise a wide project for feasibility and bankroll.

Conclusion
On theory, since years ago, every accredited ART center has its own IT stored repository. Very likely, most of the biomedical software approaches were different in the usage logics and in the architectural design of the ART records. Sooner or later it will be necessary to leverage the database to meet requirements of compatibility with a centralised donors-offspring registry reminiscent to the one we propose. The DART registry is a not relinquishable step for the ART centers thus the availability of an architectual design of an entity-relational structure together with the algorithm and control logic specification of the a system informatic platform can be adopted and validated for implementation projects.

In sharing an open view of the experience in IT databanking “for physicians”, we propose an open minded basics for anyone to pickup and transfer some “hints” from our theoretic model into local ART IT programmers.

All technological components of the model were conceived to give plausible solution to one or more of the critical areas of the DART such biomedical data privacy, repeated donation, identity disclosure control for all the subject involved in an pregnancy program (recipients, donors an ultimately offspring).

At the present the DART registry does not exists, yet we believe that is profitable to forerun times to prepare and anticipate “how-to” this system should be realised. Albeit our indication and formalisation of the registry model for DART is mostly based on a theoretic paradigm, yet it provides technical and functional specification for a future project and that is why we judge useful to openly disclose our research.
References


3. Salvatore A. Reina, Vito M. Reina, Eugenio A. Debbia “Records matching model for data survey on applied and experimental microbiology” (December 2007); NEW MICROBIOLOGICA, 30, 35-44, 2007


11. Kushida, Clete A. MD, PhD; Nichols, Deborah A. MS; Jadrnicek, Rik; Miller, Ric; Walsh, James K. PhD; Griffin, Kara MA - Strategies for De-identification and Anonymization of Electronic Health Record Data for Use in Multicenter Research Studies. Medical Care: July 2012 - Volume 50 - Issue 7 - p S102–S101; doi: 10.1097/MLR.0b013e3182585355


Acknowledgments

S.S.D.Medicina della Riproduzione, Ospedale Evangelico Internazionale, Genoa, Italy Group; DSCDI, DISC Faculty, School of integrated Medicine, University of Genoa : DIBRIS – Department of Informatics Bioengineering Robotics and System Engineering, CEBR – Center of Excellence for Biomedical Reasearch, University of Genoa

Software Copyrights & logos citations

The imaging works of the paper were obtained by utilising
C1. VUE (Tufts University, Boston, Medford/Somerville, Massachusetts, USA) and

Statistic trials and tests on data samples and raw records’s collections were performed by utilising
C3. PAST by Øyvind Hammer (Natural History Museum, University of Oslo, Norway),
C4. JMP (SAS Institute Inc., NC 27513-2414, USA) and
C5. ViSta by Forrest W. Young (Department of Psychology University of North Carolina CB 3270 Davie Hall, Room 341b Chapel Hill, NC 27599-3270).
C6. Partosoft (c), Semesoft (c), SemeBOX 3.0 (c), RiproSoft (c) software libraries and UML design were kindly provided by RR Associated Brains (c) internet open group according Creative Commons licence extensions.
APPENDIX

Table 1 – Major Critical Control Points in DART requirements and practice controls

<table>
<thead>
<tr>
<th>Critical Control Point</th>
<th>DART Implication</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Anonymisation of donors</td>
<td>Donor and receiving patients will not know each other and by this newborn can be considered only for health reasons. Waive to this rule will necessary to inform institution and get mandatory permission prior to exchanging and communication sensitive information what so ever.</td>
</tr>
<tr>
<td>2. Donor, gamete sample and adversity traceability</td>
<td>The traceability traverse almost all the procedures of the DART program and all the operative steps have to be monitored accordingly a common privacy validation process of Data Protection.</td>
</tr>
<tr>
<td>3. Sequential donation</td>
<td>Donors have to be associated to a center once and gamete sample can be used for more than 10 pregnancies</td>
</tr>
<tr>
<td>4. Fake identity control</td>
<td>Either, masking or substitution of the donors who impersonate diverse persons is always possible and it should be prevent by document and biometrical controls.</td>
</tr>
<tr>
<td>5. Age ranges limits</td>
<td>Donor and receiving patients will fit adequate minimum and maximum values</td>
</tr>
<tr>
<td>6. Clinical indication for DART enrols</td>
<td>Pre-assumption of the clinical indication for the reproductive techniques access have to be certificated. Applied criteria are tabulated pending on the subject and its sex.</td>
</tr>
<tr>
<td>7. Criteria validation for the access to a DART program (details in Fig 2)</td>
<td>Tests and screening for infectiological, hematological, genetic, anamnestic laboratory screening limitation have to be verified for a donors and their sample to access to the donation program.</td>
</tr>
<tr>
<td>8. Unique identification coding</td>
<td>A universal and standard coding have to be used either locally inside the Tissue Institute or at European level (EUSTITE projects). One code have to be applied to each individual subject of the DART (including biological material) and the combination of associated codes have to be classified across temporal timestamp.</td>
</tr>
<tr>
<td>9. Selection of the characteristics of donor</td>
<td>It is not possible to choose particular phenotype’s characteristics for patients. The DART centers have to assure a major level of compatibility between couple and donor phenotype characteristics.</td>
</tr>
<tr>
<td>10. Informed consent</td>
<td>Couples, patients and donors all need to pose a written signature on the official documents of the center privacy consent and notification.</td>
</tr>
<tr>
<td>11. Selection of donor profile</td>
<td>Validation and choice of a donor need to be befited by a team of geneticist, psychologist, biologist/embryologist, endocrinologist/urologist, andrologist supervised by the chief of the accredited ART Center.</td>
</tr>
</tbody>
</table>

(§) – temporary according to last guidelines from Italian “Conferenza Permanente Stato Regioni”

Table 2 – Cumulative ART indicators of intramural database (2003-2014)

<table>
<thead>
<tr>
<th>Item</th>
<th>Cycles</th>
<th>Couples</th>
</tr>
</thead>
<tbody>
<tr>
<td>I° level ART</td>
<td>1012</td>
<td>335</td>
</tr>
<tr>
<td>II° and III° level ART Cycles</td>
<td>5509</td>
<td>2543</td>
</tr>
</tbody>
</table>

Notes: in the 2011 our system underwent a major upgrade and moved to a different standard of database format. For the sake of computation in the modelling of the pattern-matching trials records between 2003 and 2011 were used (1812 couples’s recs).
Table 3 – Principal data variables grouped by logical abstraction of the ER tables in DART-KB database

<table>
<thead>
<tr>
<th>Logical group</th>
<th>Most relevant fields/variables</th>
<th>PMV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anamnesis and anthropometric profile</td>
<td>DOB, BMI, somatic phenotype, biometry, demography</td>
<td>yes</td>
</tr>
<tr>
<td>Pre-conceptional tests</td>
<td>Preliminary screening of legal exams with serology and mandatory serological markers</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(cystic fibrosis, VDRL, HIV, Toxo, Rubeo, blood type, etc)</td>
<td></td>
</tr>
<tr>
<td>Basal stimulation</td>
<td>Initial woman situation (ovary metrics, weight and height, menstrual, gynecological echography)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>etc.), E2 dosage, precocious FSH dosage</td>
<td></td>
</tr>
<tr>
<td>Stimulation Monitoring</td>
<td>Gonadotropins and antagonist therapy, timeline indicator curves, outcome stimulation, flow</td>
<td>yes</td>
</tr>
<tr>
<td></td>
<td>status if destined to IUI (1° ART level)</td>
<td></td>
</tr>
<tr>
<td>Pick-Up phase</td>
<td>Operator, complication, follicles, bleeding, accessibility, Vescical lesions, Hydro- salpinx,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fibrome, destinate and not destinate oocytes, oocyte source, rescue metrics</td>
<td></td>
</tr>
<tr>
<td>Embryological evolution</td>
<td>Precooling classification, freezing / thawing details, Hyaluronidase incubation, medium and</td>
<td></td>
</tr>
<tr>
<td></td>
<td>cultural, ART technique, conditions, incubation timing, reinoculation, type of fertilization,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>blastomers morphological evolution, fresh vs thawed embryous usage, transfer /GIFT</td>
<td></td>
</tr>
<tr>
<td>Outcome followup</td>
<td>Tracking of Pregnancy and Newborns outcomes</td>
<td>yes</td>
</tr>
</tbody>
</table>

\) The PMV column means that corresponding fields/variable lists where used in the Pattern Matching Engine as well as for the Inferential Engine in the DART-KB of the Registry model. Please note that variables are cited only as descriptive examples.
Figure 1 – Clinical indication to the DART

Clinical Indications

- **Feminine**
  - Age: 20-35 years old
  - Spontaneous
  - Selfless
  - On going / past ART enrols

- **Masculine**
  - Age: 18-40 years old
  - Spontaneous
  - Selfless

- **Patients**
  - iatrogenic and medical situation of proved sterility for which oocytes are not available and/or usable

- **Donors**
  - iatrogenic and medical situation of proved sterility for which sperm samples are not available and/or usable

**Heterologous Fecundation**
**Figure 2** – Cascade flow of the Donors selection criteria

Principal features: good health status, no genetic anomalies across the family, sibling and genetic anamnesis (questionnaire). Donor provides clear and complete information of his biological parents and formerly he was not adopted, nor himself conceived with ART.

Donor have to be judged capable of discernment with no evidences nor suspected probabilities of inhereditary disease in the family tree.

Donor should be considered to exclude economical or emotional reasons that can compromise the correct disposition to the donation.

- Selection of donors with proved fertility is desirable not enforced
- Psychological counselling is recommended for all donors with interview and appropriated tests
- No ART Center operators, owners nor personnel can be a donor
- No patient’s physician who execute the ART can be a donor
- Men who belong to high risk category because of professional exposure to toxic substances can not be donors
- Persons who underwent chemotherapy or radiotherapy later than two years can not be gamete donors
**Figure 3** – ER Cascade of the ART logical design of the scheme in Table 3

Note: GIFT technique is currently obsolete, nevertheless it was maintained in the model because of records retro-compatibility. In our case the database was initiated in the 2003 and we did not want to exclude the possibility of retrospective analysis of historical data.
Figure 4 – Automated Inferential flow-chart of the CCP across ART process

Critical Control Points are innerly and transparently handled by the heuristic engine as the interaction with physician proceed ahead to dynamically prepare subsequent steps.
Figure 6 – Overall landscape of DART-KB platform model and Cloud Computing layers interaction