

# Suggested Approach Using Cloud Computing And DNA Test in Finding Missing Children In All Over The World

Suhad Abu Shehab  
A Scientific Research  
suhad\_cis@hotmail.com  
Facility Of Information Technology

Dr. Mostafa Alrawashdeh  
(Lawyer, Author and Cyber Crimes Expert)  
mostafaalrawashdeh1155@gmail.com  
Facility Of Law

**Abstract** - Wars, natural disasters, conflicts in some areas, abandoned and kidnapped children .. etc. all these juveniles cause a big human's problem all over the world (unknown and foundlings Childs). Parents and relatives to those unknown children try everything to find their children, it became an impossible mission in some situations like wars, natural disasters .. etc, because of the huge numbers of unknown children they be founded by other people, rescue teams, and any other way especially when child moves from country to another country or parents are forced to leave their countries. To solve this huge human's problem, countries must use law and technology of database and comparisons algorithms between the DNA of the child and DNA of their parents or relatives.

**Key Words:** Cloud computing, Unknown Children, DNA test , Data Mining, Algorithms, child's rights, law.

## I. INTRODUCTION

War, Natural disasters and other juveniles caused a lot of families to lose children's during these situations, they try as possible as they can to find their children's, but they arrive to an end road without finding their child , his or her child become a missing child. "A missing child or unknown child is a child whose whereabouts are not known to the parents, legal guardian or any other person or institution legally entrusted with the custody of the child, whatever may be the circumstances or causes of disappearance, and shall be considered missing and in need of care and protection until located or his safety and well-being established" [1]. In the other hand these juveniles caused to find a lot of very young children without knowing anything about their parents , their names or anything about them, that makes finding their parents during this mess an impossible mission. The same thing about foundling's child, "Foundling' is an historic term applied to children, usually babies, that have been abandoned by parents and discovered and cared for by others" [2]. In this case we cannot know anything about the child or about his or her parents. It becomes so hard in the future to know anything about his or her root. In this paper we will call foundling, missing and unknown children the (Unknown Children).

The solution for this painful problem is a combination of all human's knowledge from different felids with each other (Information technology, biology and law). In this paper we come up with a suggested solution for this problem by suggestion of establishing a huge DNA cloud database for all children without knowing their parents or anything about them in all countries all over the word, then uploading this data for international shared Cloud (Cloud computing). Every parent tries to find their

missing child's by making a DNA test then the results are uploaded and compared with the data in the cloud that has all the DNA test for unknown children to find match in it. The result of the match process is his or her missing child by the international law.

## II. USING DNA TEST TO FIND BIOLOGICAL PARENTS AND RELATIVE'S BACKGROUND

As shown in figure (1) Google search for "DNA in finding parents", a lot of websites concern about finding parents by comparing DNA test results, such as "Connecting with Your Biological Family" through DNA Test, in this web site they focus on Adoptees and others with unknown parentage they use DNA testing to find and connect with their biological families or to learn more about where their ancestors came from. In the beginning, they take a DNA Test and review of DNA matches, using some companies database to provide a DNA match list. In the next step of this process is reviewing the DNA matches. The final step is reaching Out but in this step they tell the person who tries to find his or her parents to take care of

1. *Your DNA match may not know how to help you determine your birth parents or immediate family.*
2. *Your birth and subsequent adoption may have been kept a secret from other members of the birth parents' families.*
3. *Your birth family may not wish to make a connection*" [11].

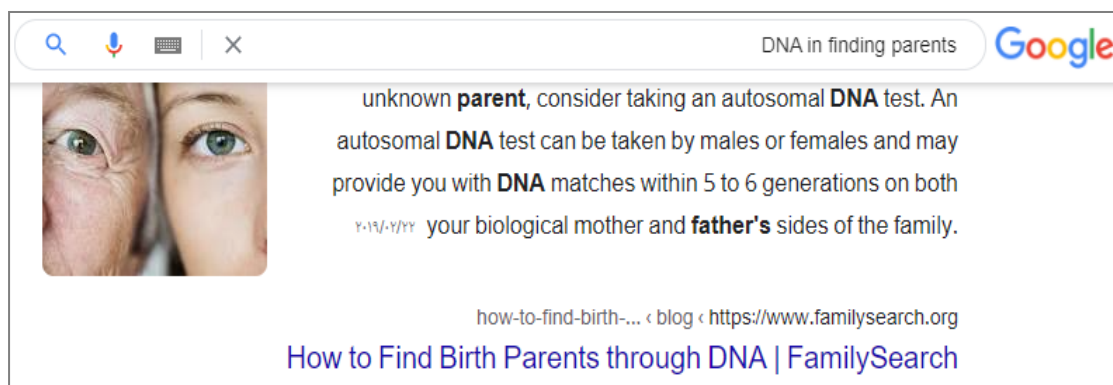


Figure (1) Google search for "DNA in finding parents" [9]

As shown in figure (2) Google search for "DNA in finding relatives", a lot of websites concern about finding relatives by comparing DNA test results, such as "MyTrueAncestry - Compare your DNA test results , Genealogy search - Free Ancestry Database and Finding Biological Family - Ancestry Support .. etc". All these websites concern to find matches by comparing the DNA results to find relatives. these websites are free and they are just for people in the UK , they are an online access to family history records. All their teams are volunteers to create high-quality transcriptions of public records from governmental sources, parish churches, and other trusted institutions. They have an *Open Data* and *Open Source* to make and keep public records accessible to all. The databases which are created freely, are available for people to search in order to support their family history research [12].

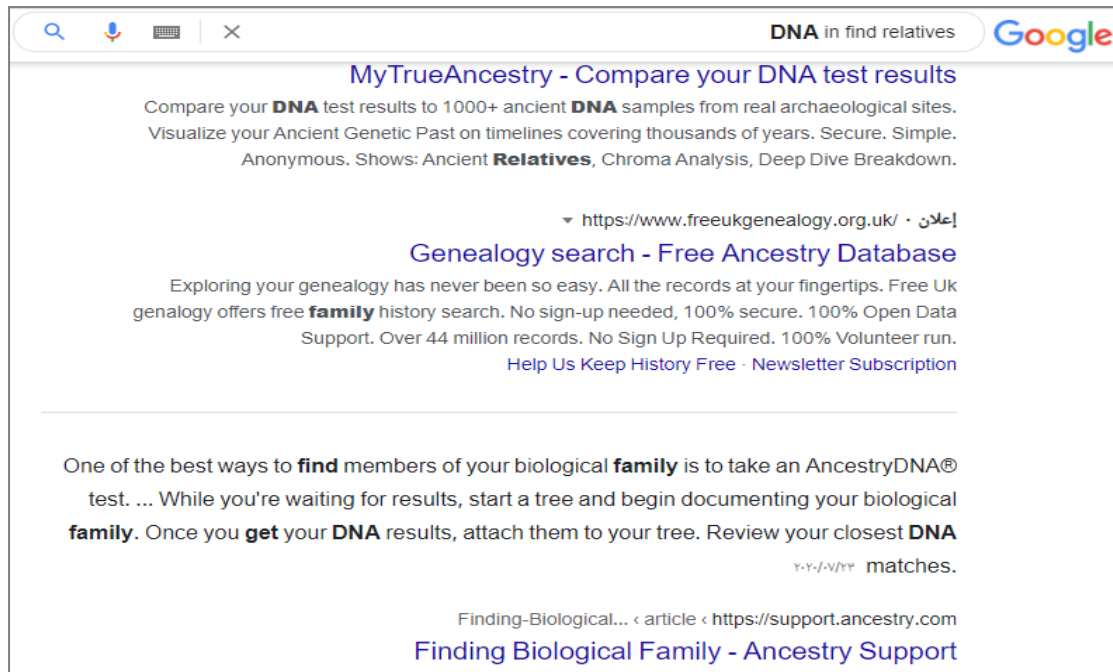


Figure (2) Google searches for " DNA in finding relatives" [10]

In figure (3) Google search for " Unknown children in cloud computing" as seen there is no papers or websites about establishing a database or cloud for all unknown children in the world.

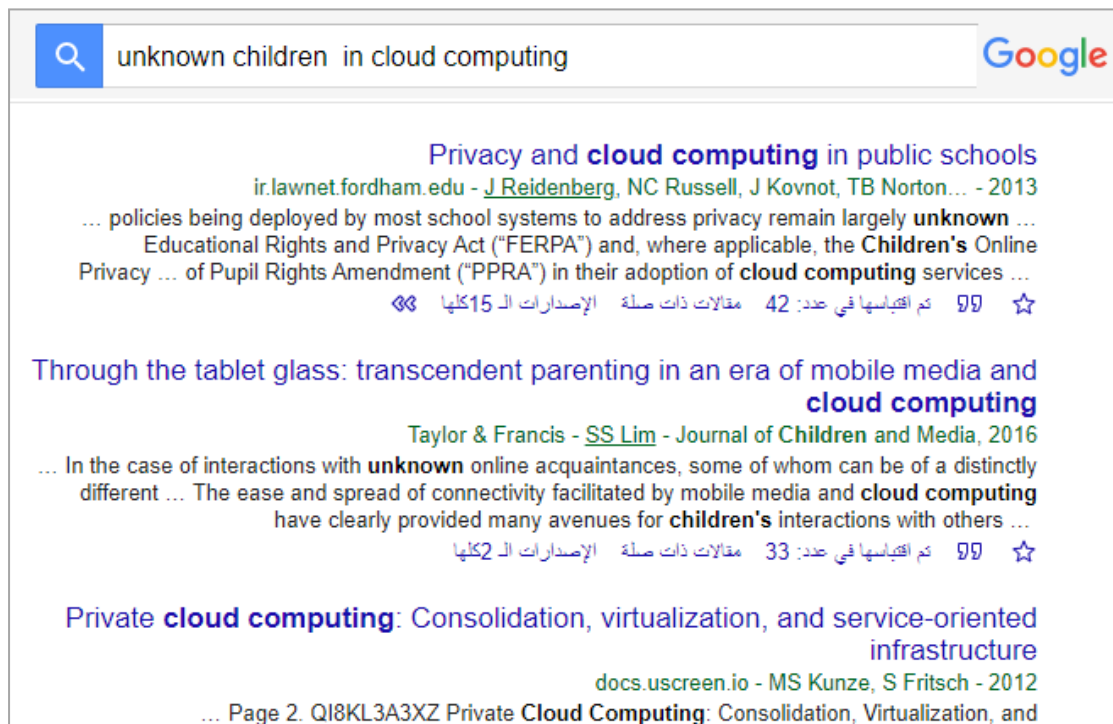


Figure (3) Google search for " Unknown children in cloud computing"[14]

These websites concern about adults trying to find their biological parents or relatives. Some of these websites charge money for their service and some of those are free of charge. The person who tries to find his or her parent must enter his or her DNA test to find the match from other DNA results

uploaded from other adults. These websites are working in some countries only. In this paper the most concern issue is establishing a formal shared global cloud database for all unknown children's without parents or relatives, using the power of technology and algorithms to find their biological parents or relatives in all over the world. In this suggested approach the most important thing is children's reunion to their parents.

### III. UNKNOWN CHILDREN PROBLEM AND CLOUD COMPUTING

#### *A. The Problem*

Family and children's rights "The International Convention on the Rights of the Child gives all children the right to a family. The right to a family allows children to be connected to their history, and it offers a protective perimeter against violation of their rights. Children separated from their families become easy victims of violence, exploitation, trafficking, discrimination and all other types of abuse"[19]. The most painful of the human's problem when parents lose their child because of wars, natural disaster .. etc, his or her child is alone in this hard world, they may face danger in his or her own live, they have no future, living in non human conditions in some cases (in some refugee camp) and a lot of dangers we cannot think about it. The child grows up without knowing anything about himself or herself ( name, family, roots ... etc). This is the most painful issue in this modern world because of a lot of wars and disasters which happened in the last two decades. They cause a millions of unknown children who do not know anything about their families. At the same time, there are wounded parents or relatives trying to find their missing child using any way they can.

#### *B. Suggested solution Using DNA test results and Cloud Computing and the Law to solve this problem:*

In this suggested approach the solution to solve this human's painful problem is that every country in the world must involve in this solution, which is responsible for their duties in helping the parents and child's to find each other. Each country must make a DNA test for every unknown child whom are found without their parents. The DNA test is reliable because each person's genetic fingerprint is unique. The DNA test is taken by the law to proof the paternity for the child.

The only method that can help in this situation is DNA test. Each person's genetic fingerprint is unique, so we can depend on this kind of test. A biological child has the same DNA with his biological father and his or her biological mother. A DNA paternity test compares a DNA sample from an alleged father or mother and a DNA sample from a child to determine if this child is biological son for them. [3]

Part of the Paternity law refers to law underlying legal relationship between a father and his or her biological children (Proofed by DNA test) and deals with the rights and obligations of parent (father) and child to each other. Child's paternity may be relevant in relation to issues rights to a putative father's title, as well as the biological father's rights to child custody [4].

Cloud computing is on-demand availability of computer system especially data storage (cloud storage) The term is generally used to describe data centers available to many users over the Internet

(over all the world). Large clouds, often have functions distributed over multiple locations from central servers.[5]

A cloud database is a database service built and accessed through a cloud platform[13]. cloud database acts like traditional database with the added flexibility of cloud computing. Users can use it by install the software to establish a cloud infrastructure to implement the database. The key features of cloud computing is a database service built and accessed through a cloud platform ,users can host databases without the need to have a hardware equipment, support relational databases like MySQL and NoSQL databases. Accessed through a web interface or vendor-provided API. [13]

The most important features of cloud computing are **Easy of access** Users can access cloud databases from virtually anywhere, using a vendor's API or web interface. **Scalability** organizations only pay for what they use. **Disaster recovery** in the event of a natural disaster, equipment failure or power outage, data is kept secure through backups on remote servers [13].

Cloud Computing is the best solution to establish an international cloud database for creating a DNA test result for all these unknown children without parents or legal garden, because any user in the world can access and use this database if this user have the privilege to access and use it.

### C. DNA test and match algorithm

DNA is an inherited concept of inheritance from a common ancestor. Humans have 22 pairs of chromosomes in which one chromosome is inherited from the father and one from the mother, each child inherits an equal amount of DNA (50%) from the mother and the father , child inherits one copy of each chromosome from each parent. The child's chromosomes is a mixture of each parent's two chromosome copies. The biological process responsible for the transmission of chromosomes from parents to child in this way is what is called meiosis. The random assortment of these chromosome fragments during meiosis is called recombination. The end result is that each child's DNA is a random mixture of DNA from his or her two parents. Comparing the chromosomes of the siblings, lining them up, observing that some regions of the chromosomes have the same color (Figure (4)) in each sibling. This indicates that they have almost identical sequences of DNA at those locations on their chromosome. These locations on the chromosome are called "identical-by-descent" (IBD) because they were inherited from a common ancestor (in this case, the common ancestor is the mother or the father)[15].

Figure (4) shows that DNA is an identical-by-descent between distant cousins (C, D, E). Chromosomes of the common ancestors (A) and their children (B) are shown. The blue and red circles indicate chromosome segments that are IBD between the indicated chromosomes.

DNA is essential in finding parents and relatives, it is about several genetic analyses to help individual find, preserve, and share their family history using Identity-by-descent" (IBD) detecting "matches" from DNA by identify long chromosome segments shared by pairs of individuals.

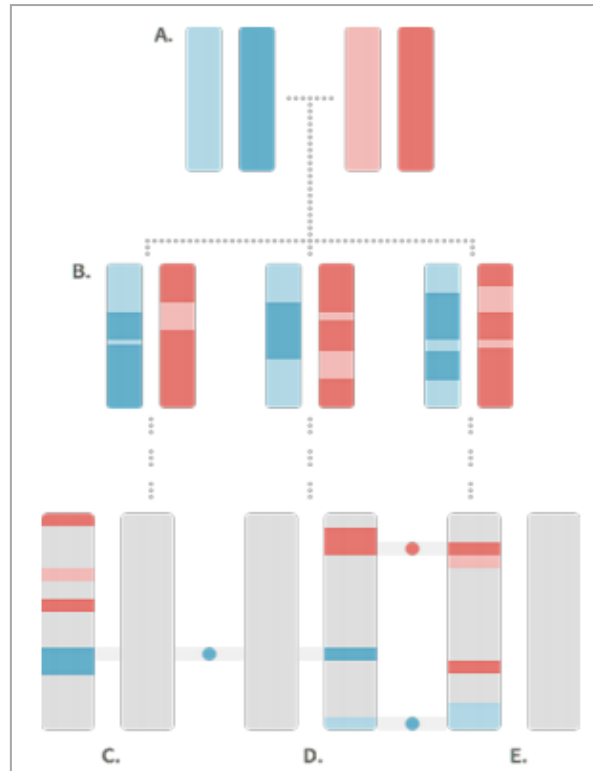


Figure (4) The process of recombination of DNA [15].

In the figure (4), all inherited some DNA from the common ancestors, only a few short segments of the chromosomes are actually identical in the same places on the chromosome of different cousins. In this example, one segment of DNA is shared by cousins C and D, and two segments are shared by cousins D and E. When IBD segments are identified, this information is used to estimate how people are related to one another by drawing connections between relatives through their DNA [15].

To improve the process of finding parents and relatives, they increased precision in identifying shared segments. The relationship between DNA matches is based on how much shared DNA is between people. In figure (5), each level of relationship is determined based on how many centimorgans matches share [16].

As seen in figure (5), when an individual knows the number of centimorgans shared with a match, it can help in understanding their relationship to them. For example, in the case of a 3rd cousin, you will share about a range of (90 - 180) centimorgans [16].

For comparison, the unit for segments of DNA is the centimorgan (cM), the human full genome is about 6500 cM. The longest length of a match is the greater chance that it is a match [17].

Approximate Amount of Shared DNA (in centimorgans [cM])	Predicted Relationship
3,475 cM	parent, child, or identical twin
2,400–2,800 cM	immediate family: full sibling
1,450–2,050 cM	close family: grandparent, aunt, uncle, half-sibling
680–1,150 cM	1st cousin, great-grandparent
200–620 cM	2nd cousin
90–180 cM	3rd cousin
20–85 cM	4th cousin
6–20 cM	distant cousin: 5th to 8th cousins

Figure (5) Relationship based on how many centimorgans matches share [16].

*D. An example of algorithm Used to find a match in DNA test*

(Multiple Bloom Filters (MBFs)): Method of storing chromosome data and searching patterns in the data, described the method of using MBFs or an array of BFs to store chromosome data in the compressed format. Subsequently, describing the procedure to search patterns in the DNA sequences that are stored in MBFs [18].

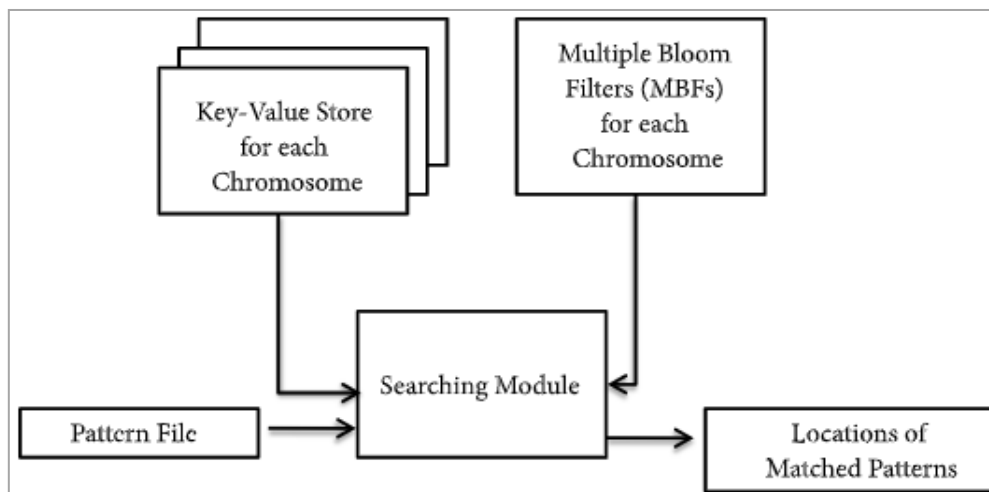


Figure (6) Multiple Bloom Filters (MBFs) [18].

This phase comprises of three important steps as seen in figure (6) the Multiple Bloom Filters (MBFs) phases: (1) Formation of k-mers of the given chromosome. (2) Storing locations of each unique k-mer in the Key-Value (KV) store on disk. (3) Storing location data for each unique k-mer of chromosome in a separate BF [18].

In the beginning of this process determined the length ( $k$ ) of DNA words, DNA is a sequence of characters called word, example (A,C,G,T) of length  $k$ . Increasing  $k$ -mer size increases number of unique  $k$ -mers and gets reduced in the sequence as length of the DNA words is increased. In this case,  $k$ -mer size is a critical parameter and has a different impact on the size of KV store and MBF. a Key-Value (KV) storage is required to store all the locations of each unique  $k$ -mer in the given chromosome. Value is the sequence of locations where specific  $k$ -mer is located in the chromosome, The next step is the construction of Multiple Bloom Filters (MBFs) for each single chromosome. Bloom filter (BF), a probabilistic data structure, is used to store all the location data of each unique  $k$ -mer in a separated BF. As a result existed a BF for each distinctive  $k$ -mer present in the chromosome. This consequently leads to the formation of MBFs, the size of the BF plays a significant role in achieving compression to determine the accurate size of the BF a pattern is a sequence of DNA characters (A,C,G,T) that is to be searched in a DNA sequence or chromosome. Before the pattern matching process starts, a pattern is decomposed into  $k$ -mers of size  $k$ . DNA word or  $k$ -mer size is set to 4 when KV store is built. The 4 is selected as the final value of  $k$  and a pattern is decomposed into  $(N-k+1)$   $k$ -mers where  $N$  is the length of the pattern sequence. After that one  $k$ -mer is part of the pattern is selected such that it has the least number of occurrences in the chromosome (target string to be searched) among all other  $k$ -mers that are part of a pattern. The next step is to select the  $k$ -mer from the pattern with the least number of occurrences in the sequence. In case, there is a tie between two or more  $k$ -mers then the one that comes first in the sorted list of  $k$ -mers is selected. KV store is an integral part which stores all the locations of each unique  $k$ -mer present in the chromosome. [18]

#### *E. Data mining techniques to manage Cloud Computing*

When we establish a database in a cloud for all DNA test results of unknown children, Data Mining algorithms are used to organize the data in cloud computing to find hidden pattern between the data, it may find relatives children (cousin children) in the cloud. Data Mining is a technique used to extract a new knowledge from existing data and emerged as a significant technology for gaining knowledge from vast quantities of data. Data mining is used to find hidden patterns "relations" between data in huge datasets, it is one of the business intelligence techniques because it can extract valuable knowledge from huge databases and find hidden relations between data [8]. The Data mining techniques is: **Classification Trees** classify a categorical variable which is dependent and based on the measurements, it divides the data nodes into small subgroup [6]. **Logistic Regression** Statistical technique that; expands the conception to work with classification Provides a method that is used to predict the probability of the occurrence of the independent variables as a function. **Neural Networks;** a software algorithm that is a set of comparable architecture of brains, network elements, weight is assigned to each unit. Different varieties of data are inputted to the input node, used a technique of trial and error, various algorithms are used to adjust the weights until it meets a certain halting criterion. **Clustering Techniques;** a technique that is used to partition a set of data into meaningful subclasses which is known as cluster. The nearest neighbour technique used finds out the gap between data. After finding that gap, it assigns the particular record to the class which is the closest neighbour in the whole data set. **Association Rule Mining;** observes that association rule mining technique involves the use of



machine learning models to analyse data for patterns in a database. Identifying the frequent and associations which are called association rules. **Machine Learning**; software can be included and learned from the data and the main focus is on making predictions.[7]

These algorithms will help in finding a match in this cloud or hidden pattern in the results of DNA test to find the relatives of the unknown child's or reunion brothers and sisters in unknown children. We can use any other algorithms or tools to find the match between unknown children and the family who tries to find his or her missing child.

#### IV .CLOUD COMPUTING IN PROOFING PATERNITY OF UNKNOWN CHILDREN USING DNA TEST RESULTS

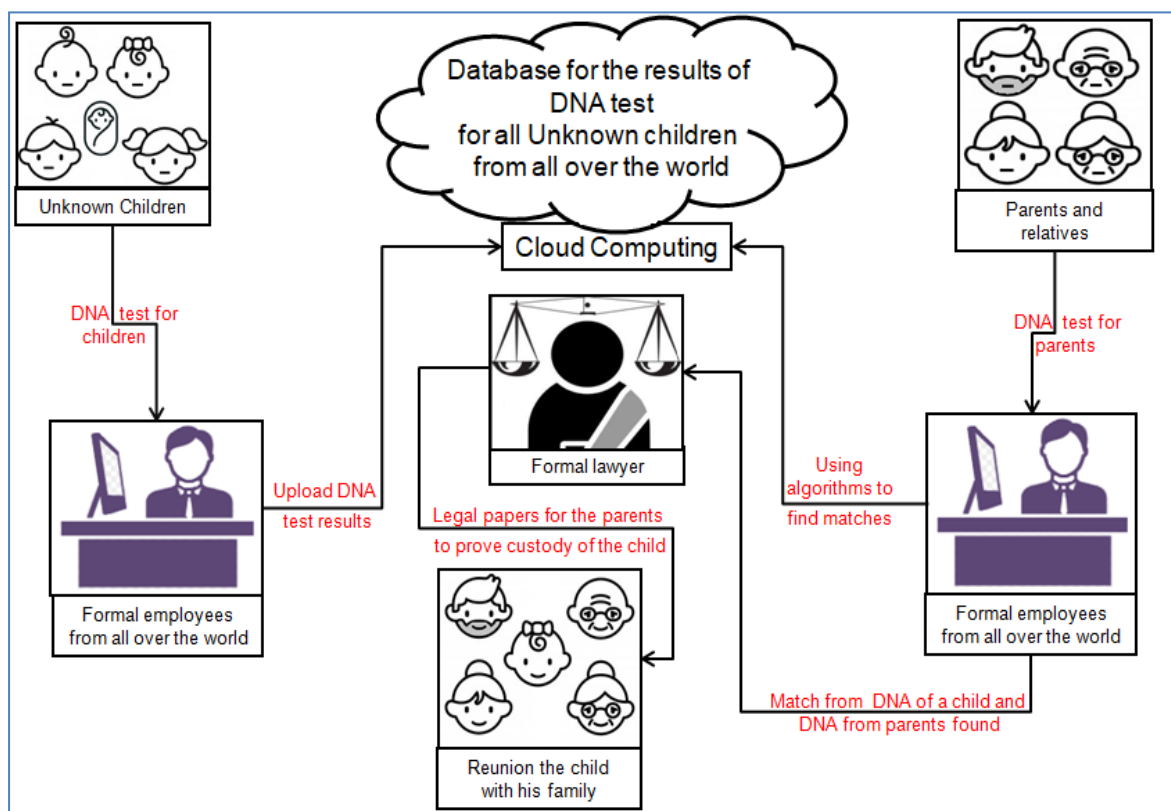


Figure (7) Cloud Computing in Proofing paternity of unknown Children use DNA test results

In figure (7) Cloud Computing in Proofing paternity of unknown Children use DNA test results, it shows the steps of the process of parents finding their missing child. For example, during a war a family lost his or her child, The family is forced to move to another country, the child's is found by rescue team. The formal employee makes a DNA test to the unknown child, then uploads the result to an unknown children database cloud.

The parents start searching of their baby, they make a DNA test and give the result to the formal employee in their country they live in. The employee use algorithm to find a match from the international unknown children database cloud. The match is found, the employee sends the result of

finding a match and the parents to the formal lawyer to communicate between countries and to make all the papers to proof the paternity of their own child. The formal lawyer makes all the procedures for the reunion process by the international law. The parent's reunion with their missing child is completed.

#### CONCLUSION

A million of unknown children in the worlds and a million of families all over the world are missing their children. If we apply this suggested approach in this paper; a thousand or millions of unknown children will reunion with their families, then one of Child's Rights will be achieved. All countries in the world must involve in the process of establishing a cloud computing database containing all unknown children DNA tests, and establishing a formal process containing all fields of (Information technology, biology and law) to make the process of reunion the child and the family possible. These children are a part of future. We must take care of them, to have a bright future of humanity.

## REFERENCES

- [1] Juvenile Justice , "Care and Protection of Children". Act, Article 92 (G.S.R. 898) , 2016.
- [2] <https://foundlingmuseum.org.uk/about/our-history/what-is-a-foundling/>, [Last Accessed May, 2021].
- [3] <https://dnacenter.com/blog/everything-you-need-to-know-about-a-dna-paternity-test/>, [Last Accessed May, 2021].
- [4] Richards, Edward P. "The Presumption of Legitimacy". Law and the Physician. The Law, Science & Public Health Law Site, 2017.
- [5] "An Introduction to Dew Computing: Definition, Concept and Implications - IEEE Journals & Magazine". doi:10.1109/ACCESS.2017.2775042. S2CID 3324933, 2017.
- [6] J.Nageswara Rao, M.Ramesh, "A Review on Data Mining & Big Data, Machine Learning Techniques" , International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7 Issue-6S2, pp 914-916, 2019.
- [7] RItu Ratra , Preeti Gulia "Big Data Tools and Techniques: A Roadmap for Predictive Analytics", International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-9 Issue-2, 2019.
- [8] Alaa H Al-Hamami, Suhad Abu Shehab " An Approach for Preserving Privacy and Knowledge In Data Mining Applications", Vol. 4, No. 1 SSN 2079-8407 Journal of Emerging Trends in Computing and Information Sciences , CIS Journal, 2013.
- [9] <https://www.google.com/search?q=DNA+in+finding+parents&ei=>, [Last Accessed June, 2021].
- [10] <https://www.google.com/search?q=dna+in+find+relatives&source=hp&ei=> ,[Last Accessed June, 2021].
- [11] <https://www.familysearch.org/blog/en/how-to-find-birth-parents-through-dna/> ,[Last Accessed June, 2021].
- [12] <https://www.freeukgenealogy.org.uk/> ,[Last Accessed June, 2021].
- [13] <https://www.ibm.com/cloud/learn/what-is-cloud-database> ,[Last Accessed June, 2021].
- [14] [https://scholar.google.com/scholar?hl=ar&as\\_sdt=0%2C5&as\\_vis=1&q=unknown+children++in+cloud+computing&btnG=](https://scholar.google.com/scholar?hl=ar&as_sdt=0%2C5&as_vis=1&q=unknown+children++in+cloud+computing&btnG=) ,[Last Accessed June, 2021].
- [15] <https://www.ancestry.com/dna/resource/whitePaper/AncestryDNA-Matching-White-Paper.pdf> ,[Last Accessed June, 2021].
- [16] <https://www.ancestry.com/corporate/blog/the-science-behind-a-more-precise-dna-matching-algorithm/> ,[Last Accessed June, 2021].
- [17] [https://en.wikipedia.org/wiki/Genealogical\\_DNA\\_test](https://en.wikipedia.org/wiki/Genealogical_DNA_test) ,[Last Accessed June, 2021].
- [18] Raihan Ur Rasool and others ' Pattern Matching for DNA Sequencing Data Using Multiple Bloom Filters' , *BioMed Research International*, Volume 2019 |Article ID 7074387, 2019.
- [19] <https://www.humanium.org/en/family-and-childrensrights/> ,[Last Accessed June, 2021].