To whom it may concern

I am sending the following documents to apply for MAP-I grants.

01	Curriculum Vitae
02	Statement of Purpose
03	Secondary School Certificate
04	Higher Secondary School Certificate (HSSC)
05	Certificate of Best Student (HSSC)
06	Certificate of First Position (HSSC)
07	Bachelor of Science in Computer Science (Hons) Degree
08	Bachelor of Science in Computer Science (Hons) Transcript
09	English Proficiency Certificate
10	Graduate Assessment Test, Score Card
11	Architecture and Design of Large-Scale Software Systems, Result
12	Model Driven Software Engineering, Result
13	Passport Front Page
14	Thesis Proposal
15	SEF Receipt

Regards

ARIF UR RAHMAN

29th April, 2008

Statement of Purpose

After obtaining my Bachelor of Science (Honours) in Computer Science degree, I knew my future would be studying more for achieving the authorization for conducting research work.

"Conceiving ideas and developing systems that deliver that promise" is a personal goal that I have set for myself. When translated into actions this goal has reinforced my belief in leadership through creativity.

The boundless possibility of trying out and the instant knowledge of the outcome that stimulates one for further analysis of a rationale in question is what I find most appealing about Computer Science. Keeping up an inquisitive and explorative attitude, I believe, leads to a constant learning process. This approach adds to the already immense potential for innovation that exists in this field.

As I have been studying Computer Science for the last six years (2 in HSSC + 4 in BS-CS), I look to postgraduate study to refine my knowledge and skills in my areas of interest. I believe it will also serve to give direction to my goal, a career as a research professional at an academic or commercial, research-oriented organization. I intend to pursue PhD degree in order to reach that goal.

In my undergraduate studies, I have benefited from the breadth of University of Malakand's syllabi content that has given me a comprehensive exposure to the core areas of Computer Science and a strong conceptual understanding of the same. In these four years of study, I have strived to maintain an approach of expending independent efforts in all my endeavors. Self learning and sharing my knowledge with others has been most worthwhile, when comprehending a concept.

At present I am in the second semester of MAP-I Doctoral program and working on Pre-Thesis under the title "Data Warehouses in the path from Databases to Archives". An American poet once observed that all history is modern history. For any given time that which is past, may have impact on our present knowledge and our future actions. Think of the impact of loosing the documents most important to someone i.e. educational / experience certificates and degrees, videos of young born babies, photos of past generations etc. Now, imagine about loosing knowledge about an entire society and the impact that would make on a nation i.e. loosing record of Tax / Social Security. It would be a tragedy of epic proportion. As all the public and

private organizations now a days rely on digital data. This data cannot be recovered if lost, or at an unbearable cost. We need to take every possible measure to preserve the data and information for the desired period of time otherwise we will have to face irreparable loss. Off course, this is a challenging area and needs a lot of work, but my PhD will be a step to preserve the past and present for future generations. After successful completion of my PhD, I see a bright future in public as well as private sector.

I am working as a researcher in the project "Data Warehouses for the Long Term Preservation of Institutional Electronic Records and Databases" at INESC, Porto. Through this work I got a good conceptual understanding of the problem. Also, this gives me an advantage in pursuing my research in the same field.

Sir, surely I would not like to over estimate myself but let me state that I believe that there are no shortcuts in life and the only road to success is hard work. I stood first in the college in the exam of Higher Secondary School Certificate (HSSC) and was acknowledged the BEST STUDENT (Copies enclosed) of the session (2000-2002). Also, qualified, Graduate Assessment Test, which is a basic requirement in Pakistan to pursue higher studies. Moreover, there has been no gap throughout my academic career.

I would be grateful to you if I'm accorded the opportunity to pursue my studies on MAP-i grants. I will try my level best to justify your faith in me.

ARIF UR RAHMAN
badwanpk@gmail.com
29th April, 2008

Curriculum Vitae

ARIF UR RAHMAN

Address: R. DO Carvalhido 155

4250-102, Porto

Portugal

Date of Birth: 15th April, 1983

Cell: + 351 916 052 435

Email: badwanpk@gmail.com

Nationality: Pakistani

Marital Status: Single

Passport No. BK1985371

NIC: 153076-666537-3

OBJECTIVE

I am fond of experimenting my own ideas. I have always struggled to maintain an approach of expending independent efforts in my entire endeavor. In any field one can only experiment his own ideas and thoughts if he has gained the maximum knowledge of the same field. A Ph.D. degree is considered a terminal degree. For my aim i.e. to experiment my own ideas, I have to acquire the Ph. D degree. I would like to make a career in the field of Computer Science as it is one of the fastest growing segments of industry, but it is also one of the fastest changing areas technologically. This field has more space for applying ones ideas and thoughts. Moreover, I have been studying Computer Science for the last six years (2 Years HSSC + 4 Years BS-CS). Therefore, I have to achieve the highest degree in this dynamic field.

EDUCATIONAL QUALIFICATION

- **Ph. D in Computer Science**

  October, 2007 - Continue...

  Theme: Data Warehouses in the Path from Databases to Archives

  Subjects Studied (First Semester)

  - Model Driven Software Engineering (Score: 14)
  - Architecture and Design of Large Scale Software Systems (Score: 13)
  - Program Semantics, Verification, and Construction (Result Awaiting)

  (www.map.edu.pt/i/2007)

- **BS in Computer Science (Hons)**

  (Nov 2002- Feb 2007) (Marks Obtained 78.17%)

  From University of Malakand, Chakdara, NWFP, Pakistan

  (www.uom.edu.pk)

- **Intermediate in Computer Science**

  (Sep 2000- Sep 2002) (Marks Obtained 69.27%)

  From Petroman Training Institute, Islamabad (FBISE), Pakistan

  (www.fbise.edu.pk)

- **Matriculation**

  (Marks Obtained 65.76%)

  From Government High School, Badwan (BISE SWAT), Pakistan
  (www.bisess.edu.pk)

SCIENTIFIC PROJECT

BS in Computer Science (Hons)

In the Bachelors of Science (Hons) in Computer Science degree course from University of Malakand, I did a project on "Appointment Management System", under a Company i.e. Sehat Consultancy, Blue Area Islamabad, Pakistan. After completion of the project, I also presented the thesis to the Examination Section, University of Malakand for partial fulfillment of my degree.

The software helps in getting complete data of an employee through a questionnaire over the web that ultimately reflects the complete history of the individual employee. This data is utilized for the well being of the employees working in certain firms / organizations such as Nokia and Mobilink. It is aimed to manage the records of medical examinations of employees which are registered with Sehat Consultancy. The Software has facilitated the Company in understanding individual health problems of the employees and has brought more business for the Company. On the other hand, the coverage of the Company i.e. Sehat Consultancy has increased three fold due to this Software.
(http://www.sehatconsultancy.com/)

EXPERIENCE

Research

(19 Nov 2007 – Continue…)

- Working as a Researcher in the project "Data Warehouses for the Long Term Preservation of Institutional Electronic Records and Databases" at INESC, Porto.

ACHIEVEMENTS

- Qualified, Graduate Assessment Test, which is requirement for higher education, in Pakistan (www.nts.org.pk)
- Got first position in College in my Higher Secondary School Certificate
- Was acknowledged the BEST STUDENT of the session (2000-2002) in College.

LANGUAGES

- **English**: Can read, write and speak frequently.
- **Urdu**: Can read, write and speak frequently.
- **Pashto**: Though my mother tongue but can only speak frequently.
- **Portuguese**: Only Basics (But am attending lectures and soon will master it)

Computer Programming Skill

C#.Net, SQL/Server, C/C++, Visual C++

HOBBIES

Internet surfing, meeting different people, visiting different places and enjoying the natural scenes.

REFERENCES

1. Prof. Dr. Gabriel David
Department of Informatics Engineering
Faculty of Engineering, University of Porto
Portugal
Email: gtd@fe.up.pt

2. Prof. Dr. Luis S. Barbosa
Department of Informatics
University of Minho, Portugal
www.di.uminho.pt/~lsb
Email: lsb@di.uminho.pt

3. Mr. M. Zahid Khan
Department of Computer Science
University of Malakand, Pakistan
Email: mzainpk@gmail.com

S.No 3433

Roll No. 34094

68

MARKS IMPROVED

BOARD OF INTERMEDIATE AND SECONDARY EDUCATION

بسم الله الرحمن الرحيم

Saidu Sharif, Swat N.W.F.P. Pakistan

Secondary School Certificate Examination

SESSION 2000 (ANNUAL/SUPPLEMENTARY)

THIS IS TO CERTIFY THAT _ARIFUR RAHMAN_

Son/Daughter of _MUSHTAQUR RAHMAN_

and a student of _DISTT: DIR ._

has passed the *Secondary School Certificate Examination*

of the Board of Intermediate and Secondary Education, Saidu Sharif, Swat held in

2000 as a *Regular/Private candidate.* He/She obtained _559_ Marks out

of 850 and has been placed in Grade B Representing _VERY GOOD_

The candidate passed in the following subjects:

1. English   3. Islamiyat   5. MATHS   7. PHYSICS

2. Urdu   4. Pakistan Studies 6. CHEMISTRY   8. BIOLOGY

Date of birth according to admission form is _FIFTEENTH APRIL_

one thousand nine hundred and _EIGHTY THREE_ (15-04-1983)

Asstt. Secretary

Secretary

This certificate is issued without alteration or erasure.

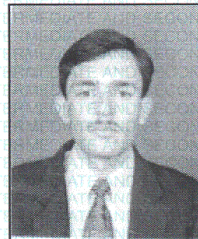# FEDERAL BOARD OF INTERMEDIATE AND SECONDARY EDUCATION, ISLAMABAD

HIGHER SECONDARY SCHOOL CERTIFICATE EXAMINATION PART- II

ROLL NO   572912

GROUP   SCIENCE GENERAL

REGISTRATION NO

CERTIFICATE NO   249022

RESULT CARD

ANNUAL EXAMINATION 2002

ARIF UR RAHMAN   Son/Daughter of   MUSHTAQ UR RAHMAN

of (Institution)   PETROMAN TRAINING INSTITUTE ALI PLAZA BLUE AREA IBD.

has secured the marks shown against each subject

in the Higher Secondary School Certificate Examination Part-II held in the month of   MAY/JUNE

Mark of Identification   MOLE ON CHIN

S.NO	Subject (s)	MARKS OBTAINED						Pass /	Remarks
		HSSC-I			HSSC-II			Fail	
		Th	Pr	Total	Th	Pr	Total		
01	ENGLISH	051		051	053		053	PASS	
02	URDU	060		060	069		069	PASS	
03	PAKISTAN STUDIES				044		044	PASS	
04	ISLAMIC EDUCATION	042		042				PASS	
05	STATISTICS	062	022	084	050	012	062	PASS	
06	MATHEMATICS	071		071	088		088	PASS	
07	COMPUTER SCIENCE	043	016	059	051	028	079	PASS	

Total Marks (In figures)   762 / 1100

(In words)   SEVEN HUNDRED AND SIXTY TWO OUT OF ELEVEN HUNDRED

THE CANDIDATE HAS PASSED AND PLACED IN GRADE B
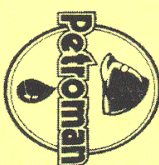
General Remarks   HE/SHE HAS PASSED THE EXAMINATION IN SINGLE ATTEMPT

Islamabad Dated: 24 - 08 - 2002   CONTROLLER OF EXAMINATIONS
Note: Error/Omission excepted   (HSSC)

"In the name of Allah the most Merciful and Gracious"

Certificate of Appreciation

Is awarded to

MR. ARIF UR REHMAN

In recognition of

THE BEST STUDENT

Of the Session 2000-2002

In the Department of ICS(Intermediate in Computer Studies)

Petroman Training Institute

Ministry of Science and Technology

Islamabad, Pakistan

Principal (ICS)

Principal (Incharge)
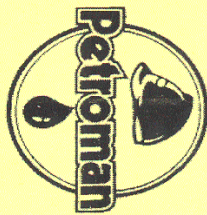
"In the name of Allah the most Merciful and Gracious"

Petroman Training Institute

Ministry of Science & Technology

Intermediate in Computer Studies Department

Islamabad, Pakistan

CERTIFICATE

This is to certify that

Mr. ARIF UR REHMAN

has obtained First/Second/Third position in FBISE Examination (2001)

We pray for his success in future life.

Principal (ICS)

Principal (Incharge)

UNIVERSITY OF MALAKAND
PAKISTAN

Serial No __BS /CS /0051__

This Degree of

Bachelor of Science (Honours)

In

Computer Science

Is Awarded to

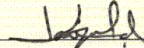Mr/Ms __ARIF UR RAHMAN__ Son/Daughter of __MUSHTAQ UR RAHMAN__

Student of __DEPARTMENT OF COMPUTER SCIENCE & I.T__

Having passed the prescribed examination held in __FEBRUARY, 2007__
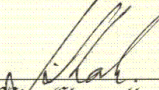
Session __2002-2006__ Registration No __2002010018__ Roll No __18__

Division __FIRST__

Issuance Date __MAY 29, 2007__

Countersigned

_____
Controller of Examinations

_____
Registrar

_____
Vice Chancellor

UNIVERSITY OF MALAKAND
PAKISTAN
DETAILED MARKS CERTIFICATE

Department of Computer Science (BCS)

Session (2002-2006)

Name: **ARIF UR RAHMAN**

F/Name: **MUSHTAQ UR RAHMAN**

BS. (Hons.) Computer Science 4 Years

Reg. No: **20020010018**

Subjects	Maximum Marks	Credit Hours	Internal	Theory	Practical	Total
			Marks Obtained			
Functional English	50	2		36		36
Islamiyat	50	2		38		38
Fundamental of Computer	100	3	18	46	17	81
Calculus-1	100	3	20	58		78
Statistics	100	3	17	59		76
1st Term Examination July 2003 Roll No: 18 — Total Marks: **400** — Result Date: Sep 08, 2003 — Obtained Marks: **309**						
Physics	100	3	19	44		63
Programming Concepts	100	4	18	52	17	87
Discrete Maths	100	4	19	69		88
Calculus-II	100	3	17	70		87
2nd Term Examination December 2003 Roll No: 18 — Total Marks: **400** — Result Date: Mar 15, 2004 — Obtained Marks: **325**						
Digital Logic Design	100	4	19	35	17	71
Data Structure	100	4	19	37	19	75
Business Communication	50	3		38		38
Programming Language-1 (C++)	100	4	20	43	16	79
Electronics	100	3	18	48		66
3rd Term Examination July 2004 Roll No: 18 — Total Marks: **450** — Result Date: Oct 22, 2004 — Obtained Marks: **329**						
Database-I	100	4	19	41	16	76
Computer Organization & Assembly Language	100	4	18	41	18	77
Visual C++	100	4	18	42	18	78
Operating System	100	4	17	46	16	79
4th Term Examination February 2005 Roll No: 18 — Total Marks: **400** — Result Date: April 27, 2005 — Obtained Marks: **310**						
Data Communication & Networks	100	4	19	35	18	72
Artificial Intelligence	100	4	18	45	18	81
Programming Language-II (Java)	100	4	19	47	19	85
Software Engineering-I	100	4	18	48	18	84
5th Term Examination August 2005 Roll No: 18 — Total Marks: **400** — Result Date: Oct 28, 2005 — Obtained Marks: **322**						
Pak Studies	50	2		35		35
Network Design	100	4	14	41	19	74
Computer Architecture	100	4	19	44	20	83
Automata Theory	100	3	15	46	16	77
Database- II	100	4	19	46	19	84
Numerical Analysis	100	4	19	71		90
6th Term Examination February 2006 Roll No: 18 — Total Marks: **550** — Result Date: April 22, 2006 — Obtained Marks: **443**						
Object Oriented Analysis & Design	100	3	17	31	16	64
Software Engineering -II	100	4	18	38	18	74
Design & Analysis of Algorithms	100	3	18	40	17	75
Computer Graphics	100	3	17	41	14	72
Internet Programming	100	4	18	48	18	84
7th Term Examination July 2006 Roll No: 18 — Total Marks: **500** — Result Date: Oct 16, 2006 — Obtained Marks: **369**						
Network Strategies	100	4	18	38	18	74
Software Project Management	100	3	18	44	17	79
Compiler	100	4	19	47	19	85
Research Thesis (Software Project)	100	6		91		91
8th Term Examination February 2007 Roll No: 18 — Total Marks: **400** — Result Date: May 25, 2007 — Obtained Marks: **329**						

Total Result Status: Maximum Marks **3500** Obtained Marks **2736**

Overall Percentage : **78.17** Total Credit Hours : **132**

Errors and omissions are subject to subsequent rectification

Controller of Examinations

UNIVERSITY OF MALAKAND
N.W.F.P, PAKISTAN

Ref.No. *uom/Acad/02/69*   Date: *5-06-2007*

TO WHOM IT MAY CONCERN

It to certify that according to the University of Malakand Statutes / Regulations pertaining to BS (Hon) / BA (Hon) Programme, English is the medium of instructions and examinations in all disciplines other than languages / Islamic studies.

This certificate is issued to Mr. Arif-ur-Rahman S/O Mr. Mushtaq-ur-Rahman on his own request.

Director Academic
University of Malakand

NATIONAL TESTING SERVICE PAKISTAN

URL: www.nts.org.pk

E-mail: support@nts.org.pk

RESULT CARD
GRADUATE ASSESSMENT TEST (GAT)-GENERAL

Sr # :

Roll No.	72410339	Valid Upto:	7th April, 2008
Name	ARIF UR RAHMAN	Father's Name	MUSHTAQ UR RAHMAN
Test Date :	8th April, 2007	Issue Date :	16th April, 2007

SCORE

	English	Analytical	Quantitative	Accumulative
Raw Score	25	12	17	54
*Percentile Score	88.97	70.40	33.81	77.21

Prepared By

Verified By

Errors & omissions are accepted within fifteen days from the date of issue.

Note: This result card will remain valid for one year from the date of test.

*The percentile score is subjected to change after each GAT-General Test conducted during the validity period of this result.

				15%	65%				20%	100%
						35%	30%	30%		
Nome	**Q1**	**Q2**	**Q3**	**Qs**	**Report**	**Exam.Q1**	**Exam.Q2**	**Exam.Q3**	**Exam**	**Final**
Andre Ferreira	90%	85%	65%	88%	75%	85%	75%	90%	79%	16
Arif Rahman	75%	75%	0%	75%	65%	65%	40%	50%	50%	13
Brett Drury	80%	85%	0%	83%	75%	70%	65%	90%	71%	15
Filipe Campos	85%	85%	0%	85%	75%	70%	60%	60%	61%	15
Hugo Ferreira	90%	85%	0%	88%	80%	95%	80%	60%	75%	16
Nuno Ferreira	95%	75%	85%	90%	80%	90%	85%	70%	78%	16
Paula Monteiro	85%	95%	80%	90%	75%	70%	70%	75%	68%	15

19 de Março de 2008

Os docentes:

Ademar Aguiar, Luís Almeida, Ricardo J. Machado

	UML/OCL		CPN		Artigo							NOTA FINAL
					Origin	Estrut	Qualid	Legib	Relev	**méd**		
António Júnior	E-	7	C	14	D	C	C	D	D	**D+**	12	**11**
Arif Rahman	---		B	17	D	C	D	D	C	**D+**	12	**14**
Brett Drury	D	11	B	17	C	D	C	C	B	**C**	14	**14**
Daniela Cruz	C	14	B	17	D	C	D	B	C	**C**	14	**15**
David Pereira	D	11	B+	18	C	C	C	D	C	**C**	14	**14**
Hugo Ferreira	B-	16	B+	18	C	B	B	B	A	**B**	17	**17**
Hugo Macedo	C	14	C+	15	C	B	B	B	B	**B**	17	**16**
Jorge Matos	C	14	C	14								
José P. Oliveira	B	17	A-	19								
Miguel Silva	C	14	B+	18								
Nuno Carvalho	C-	13	B	17	D	D	D	D	D	**D**	11	**13**
Nuno Ferreira	E+	9	C+	15	C	B	B	B	B	**B**	17	**15**
Paula Monteiro	D	11	C+	15	D	C	C	D	B	**C**	14	**14**
Paulo Jesus	B+	18	A	20	C	B	B	B	C	**B-**	16	**18**
Ricardo Vilaça	C	14	A-	19	D	B	C	D	C	**C**	14	**15**

A	Excellent	20
B	Very Good	17
C	Good	14
D	Satisfactory	11
E	Insufficient	8

To whom it may concern

Arif ur Rahaman came to Portugal about a year ago with a research grant for a project in Porto University and also to attend the MAP-i lectures at his own expenses. This shows his strong character and commitment to pursue his PhD studies with success.

In my role as coordinator of MAP-i Organising Committee I had the opportunity to witness his commitment to the course and his efforts to follow the lectures and take the maximum out of it. He also did a remarkable work as my student in the Model-based Software Engineering course.

I therefore recommend Arif for a PhD grant which is probably the only way available to him to pursue his PhD studies within the Programme, in which I think he will be highly successful.

I should also add Arif is a hard-working student and a very kind and educated young man, both from a scientific and humanistic point of view.

Universidade do Minho, Braga, 29th April, 2008

Luis Soares Barbosa
Associate Professor

Logic and Formal Methods Group,
Dep. Informatics, Minho University
4700-057 Braga, Portugal

REFERENCE LETTER

TO WHOM IT MAY CONCERN

It gives me utmost pleasure to write a few words about **Mr. Arif ur Rahman**, Registration number: 200 200 100 18, who was my student from 2002 to 2006 during his BCS (4-Years Hons - Bachelor in Computer Science) from Department of Computer Science & I.T. University of Malakand, Chakdara, Pakistan. I have taught him Introduction to Networking, Network Security, and Computer Architecture. His performances were to the utmost satisfaction. He is thorough hardworking and highly dedicated.

He submitted his assignments on time. I always found him taking keen interests in his studies. I do expect that he would be a man of brilliant career. Knowing him as a student and seeming his background handwork and devotion to studies, I have no doubt to recommend him for any kind of grant and scholarship.

 I wish him best of luck and success for his brilliant future.

Muhammad Zahid Khan,

Lecturer in Computer Science & I.T
University of Malakand
www.uom.edu.pk

mzainpk@gmail.com

Ministry of Interior,
Government of Pakistan
requires and requests
in the name of
The President
Islamic Republic of Pakistan

all those to whom it may concern
to allow the bearer
to pass freely without let or hindrance
and to afford the bearer such assistance
and protection as may be necessary

Director General
Immigration & Passports

A4102285

ISLAMIC REPUBLIC OF
PAKISTAN

PASSPORT

Type	Country Code	Passport Number
P	PAK	BK1985371

Surname
ARIF UR RAHMAN

Given Names

Nationality
PAKISTANI

Date of Birth
15 APR 1983

Citizenship Number
15307-6666537-3

Sex
M

Place of Birth
DIR LOWER, PAK

Father Name
MUSHTAQ UR RAHMAN

Date of Issue
29 APR 2007

Issuing Authority
PAKISTAN

Date of Expiry
27 APR 2012

Tracking Number
10601155973

Booklet Number
A4102285

P<PAKARIF<UR<RAHMAN<<<<<<<<<<<<<<<<<<<<<<<<<<<
BK19853713PAK8304151M12042741530766665373<00

<div align="center">

PhD Thesis Proposal

Data Warehouses in the Path from Databases to Archives

Supervisors: Gabriel David, Cristina Ribeiro
Faculty of Engineering, University of Porto

</div>

Introduction

Organizations are increasingly relying on databases as the main component of their record keeping systems. However, at the same pace the amount and detail of information contained in such systems grows, also grows the concern that in a few years most of it may be lost, when the current hardware, operating systems, database management systems (DBMS) and actual applications become obsolete and turn the data repositories unreadable. The paperless office increases the risk of losing significant chunks of organizational memory and thus harming the cultural heritage.

Significant research addressing this concern has already been conducted. The conclusions discard approaches now considered naive like trying to preserve specimens of the machines, system software and applications, in all their main versions, so that the backups of every significant system could be used whenever needed. A variant of this, instead of preserving the hardware, suggests simulating the older hardware in newer machines. More promising research suggests the conversion of database contents into an open neutral format with a significant amount of semantics associated (XML dialects), so that it becomes independent of the details of the actual DBMS.

The present proposal stems from this principle but tries to go a step further based on the following observation: there is a parallel in the attitude of a data warehouse designer approaching a database-centered operational Information System (IS) to specify a Data Warehouse (DW) and an archivist analyzing a document-centered organizational IS to specify an archiving policy and system. Both search an integrated model of the organization, merging information from a diversity of sources, systems and technologies, both have a process-centric methodology, specifying data marts or classifying related series of documents, both have long-term validity and integrity requirements, both have an evaluation attitude, leaving out irrelevant details in the data or in the series of documents to concentrate on the essential, both want to build an archive which remains basically unchanged, except for the addition of newer data or documents, and both want to expose the respective information contents in a simple and systematic way.

Of course there are differences, first of all in the respective goals. The DW designer usually tries to answer the information needs of the organization management from the point of view of decision support, monitoring, trend analysis and forecast, while the archivist wants to preserve the memory of the organization and its processes, for future generations. So concrete decisions on evaluation and elimination procedures may differ, according to the specific requirements, but the general working framework seems similar.

Thesis

Following this basic intuition, the research question proposed is to explore the adequateness of the DW approach as a target vehicle to perform, with respect to a given IS, the functions considered essential from an archivist viewpoint like appraisal, classification, elimination, description, and access while respecting properties like authenticity and integrity.

Context

Archivist methods have been put under great strain by the growing amount of institutional information that is being stored in digital format. Although the goals and principles of the discipline remain solid, their application to the new supports and information structures is subject of much debate [1,2,3].

The domain of research in this proposal is archiving data records produced as a result of the regular activity of an institution, which are usually kept by a DBMS.

At first, dominant approaches to the preservation of electronic records were either too much biased to the paper-based procedures or unfeasible: the projects of preservation of both the data and the tools to read it in the original format, including hardware and the several layers of software (operating systems, DBMS, applications), in the appropriate versions, despite its value from a museum perspective, were soon recognized as an unfeasible solution to the archivists goals. The same happened to the projects betting on simulating the old machines. The third alternative is collectively labeled as migration, though it has many trends.

Two research areas are essential for this proposal. The first has focused on the description of documents to render them available for retrieval, across the frontiers of domain modeling, document nature and storage technology, and has grown along with the Web. The second has roots on the concerns of archivists with the fragility and opacity of digital materials, and has a broader research agenda fuelled by the needs of organizations and increased awareness of the need for new approaches.

From a pragmatic viewpoint, the first line has originated several standards to add metadata to current digital objects (Dublin Core in the case of generic objects, RDF for Web objects and EAD for digital archive objects) and thus contribute with partial solutions to the description problem. However, the general goal of building digital archives needed a more comprehensive treatment which resulted in the OAI standard[4]. The second line has the more ambitious goal of dealing with the general problem of preserving the wealth of information that is being generated in digital form or converted to it. Several projects have been dealing either with fundamental models for integrating preservation into the management of current records or with solutions to the concrete problems that arise in a specialized domain [5].

One of the basic solutions is to serialize the database, systematically storing the data dictionary (table names, columns, integrity constraints) and the actual values of each column in each table line. This way, one is able to record all the values of every database with a single archiving model. The main problem of this approach is that it forgets that the data is just part of the problem in a database system. Most real information systems are structured in three layers: data + business rules + presentation. If the presentation layer may contain not too much knowledge, both the data and the business rules layers keep their own part of the semantics of the data. In certain cases, the values are meaningless without the code that discloses their interpretation. The solution envisaged is to perform

a previous step of eliciting implicit knowledge in the application code and storing it as explicit columns in a new data model. This operation is a typical step in a DW design process.

The Proposed Approach

The process of specifying a DW can be used for building an information asset that is both faithful to the original data and organized in a way that can be given use in the future without the complexity of the original system. Although the DW primary intent has been to support management decisions with flexible and relevant data (a goal more typical of current archives), the tendency to add more and more data to it turned DW into the most valuable repository of large organizations. Understanding this, recent recommendations on DW design [6] stress the importance of a global, process-centred analysis of the organization, resulting in the clear establishment of a DW bus architecture where dimensions, a kind of authority files, are initially identified and then reused in the several stars that constitute the dimensional model. The data considered relevant includes all basic facts, to enable arbitrary future queries. The knowledge of techniques to deal with changing dimensions without information loss is also central to give the DW an archival value.

Research Goals

The research explores the application of DW technology to the preservation of complex electronic records.

In more detail, the research will study:

- Properties that the DW must possess, according to the standards already established (like the Open Archives Initiative);
- Rules of transformation from operational systems into DW, adopting a process-centric but integrated view, which guarantees those properties;
- An **XML version** of the DW, though the DW model is already very simple and thus fulfils the requirement of platform independence required by long-term preservation;
- **Metadata** needs, beyond the data in the operational database, to preserve the meaning of the processes, most often just implicit in the data but present in organizational procedures and in software development documentation;
- **Application** to a real system of a concrete institution, playing the role of a case study;
- **Assessment** of the results obtained with the help of external experts.

The research is expected to have impact in the area of archiving automated information systems in the same measure of the generality and simplicity of the found solution. An immediate benefit would be the migration of a complex database into an archival format. In a more general setting, the research would contribute to preserve the digital components of organizational memory for future exploration.

Data Warehouse for Archives Model

DW design methodologies were initially derived from the goal of producing decision support systems. However, the maturity of this area along with the increasing computational power available

led to more conservative design principles, based on the idea of storing all the basic facts in their finest granularity. The proposed DW bus architecture stresses a global analysis of the organization to identify in the first place all the relevant dimensions, i.e. the authority files and other entities, and then the main processes that will correspond to the facts tables. This very simple star schema, with a central fact table surrounded by a set of dimensions, is at the heart of the dimensional model and of the whole idea of DW. The goal is to organize the information in a simple and systematic model, so that optimized search and processing is possible.

In this research direction, additions to the basic model should be investigated, in particular which dimensions and attributes are mandatory or recommended from the viewpoint of an archivist, which classes of attributes are required in the fact tables, which mechanisms of evolution for dimensions are acceptable, and in general which properties must the DW possess, according to the standards already established (like the Open Archives Initiative).

A set of typical organizational situations should be gathered and example solutions to the corresponding modeling problems provided.

Transformation Process

After the definition of the target DW model and its implementation in a concrete DBMS, it is necessary to define the whole process of populating it with the data coming from the available sources. Conceptually, the central problem of this research proposal is to get data from the operational IS in the organization and transform it into the DW. At this level, the problem is a mapping between models. The conclusion of this study will give rise to a set of mapping rules.

Implementing the rules will correspond to concrete extraction, transformation and loading processes where data will be selected, perhaps eliminated, cleaned, formatted according to the standards, checked for referential integrity against dimensions fulfilling the role of authority files, and transferred. These process definitions will also be gathered.

Taking into account the experience of archivists in this task is essential as it goes through the phases of appraisal, selection, elimination and the automated part of description, typical of the archivist process. The system must support the addition of possible bits of manual description.

This task is the main step in the migration approach proposed. Data is not only subject to a technological conversion, as in many migration proposals, but also to a controlled and documented reorganization that is necessary for the survival of the information. The XML version of the data results from a second step of migration, this time to ensure portability and long time preservation of the information but performed essentially at a technological level and keeping its structure unchanged.

XML Model

Although the DW model is already very simple and thus fulfilling the requirement of platform independence required by long-term preservation, a XML version of the DW will be established, less suited to the arbitrary querying the DW provides, but more adequate to information exchange with other systems. XML is becoming a lingua franca for automatic but still human readable information exchange, and so the ability to bidirectional conversion between XML and the DW is unavoidable. Some will even prefer XML for long-term preservation for the sake of uniformity with other kinds of digital objects preservation, but with reduced ability to offer access services to the users.

Metadata Requirements

The design process of a DW typically includes the definition of metadata describing the exact process of obtaining a certain fact, the source, the loading date, the tool version used for it, etc. These elements will probably be kept in the extended dimensional model. However, a lot more contextual information will be found mandatory to add more semantics to the data. For instance, the operational system UML or E-R models may be considered important, as well as a data dictionary, explaining the meaning of each attribute in each dimension and fact. Other contextual aspects concerning the organization around the IS and its evolution, the implicit or explicit definition of processes, forms, data sheets etc. may also be important to improve the understandability of the data collected.

A second set of metadata elements will derive from the adaptation of the would be archivistic processing of the electronic records. This includes information on the corresponding appraisal, selection, elimination, and description activities. In particular, one will investigate the implications of the ISAD(G) and ISAAR(CPF) standards on the description of such records and devise ways of extracting parts of that metadata during the ETL processing. It will, for instance, explicitly state the institutional producer of a certain record, that stands for the corresponding fact, and under which mandate he is responsible for the fact.

Conclusion

To carry on the agenda of this proposal interdisciplinary knowledge must be gathered, with knowhow in the domains of archives, information systems and data modeling.

References

[1] *INTERPARES Project - Preservation Task Force Final Report*. University of British Columbia. 2001. <http://www.interpares.org>

[2] *Records Continuum Research Group*. Recordkeeping Metadata Project, http://www.sims.monash.edu.au/research/rcrg/research/spirt/about.html

[3] Digital Preservation Coalition. http://www.dpconline.org/graphics/

[4] OAI- Open Archives Initiative. http://www.openarchives.org/

[5] Margaret Hedstrom. The Digital Preservation Research Agenda. Proceedings of the Conference on The State of Digital Preservation: An International Perspective.[2002]

[6] Ralph Kimball, Margy Ross. The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling (Second Edition). John Wiley & Sons, 2002.

[7] Cristina Ribeiro, Gabriel David and Catalin Calistru. A Multimedia Database Workbench for Content and Context Retrieval. Proceedings of the 2004 IEEE International Workshop on Multimedia Signal Processing. IEEE Press 2004.

SERVIÇO DE ESTRANGEIROS E FRONTEIRAS

RECIBO COMPROVATIVO DE PEDIDO DE:

☒ Título de Residência   ☐ Prorrogação de Permanência   ☐ Igualdade de Direitos

☐ Renovação de T.R. n.º _____   ☐ Nacionalidade Portuguesa

☐ Emissão de 2.ª via   ☐ Certidão   ☒ Outros **Art. 91º-3**

Nome: **ARIF UR RAHMAN**

Data de Nascimento: **15 / 04 / 1983**

Nacionalidade: **PAQUISTANESA**

Proc. N.º: **143 08**

O Funcionário

08.01.08

Obs.: *O presente recibo é válido por 60 dias e comprova que o interessado requereu a emissão do documento acima assinado, devendo ser apresentado no acto de levantamento.*

Mod. 100