BCS HIGHER EDUCATION QUALIFICATIONS
Level 5 Diploma in IT

April 2012

EXAMINERS' REPORT

Database Systems

Question A1

Examiners Comments

This was a popular question – 85% of candidates attempted it

A1. The following report shows the money taken by films being shown at several cinemas across the UK. Each cinema has a manager who records the amount of money taken whilst the film is shown at that location.

The following table shows sample occurrences of the data which is a complete representation of the scenario:-

Fno	Fname	Cno	Cname	CLocation	Mno	Mname	Takings
15	Arthur	1	Odeon	Newcastle	01	Green	£220
		2	ABC	Newtown	01	Green	£170
		3	Embassy	Croydon	03	White	£500
45	Titanic	1	Odeon	Newcastle	01	Green	£600
		2	ABC	Cardiff	01	Green	£880
		3	Embassy	Croydon	03	White	£290
		4	Odeon	Edinburgh	04	White	£430
71	Rocky	1	Odeon	Newcastle	01	Green	£180
		5	Gaumont	Warwick	02	Wood	£125
78	Jaws	2	ABC	Cardiff	01	Green	£150
		3	Embassy	Croydon	03	White	£200
		6	Gaumont	Bristol	05	Brown	£290
88	Arthur	7	Odeon	Croydon	02	Wood	£225

(a) Identify any repeating groups in the above representation and describe how they can be removed to produce a relation in first normal form. Identify the key attributes of the resultant relation.

(5 marks)

Repeating group is Cno, Cname, CLocation, Mno, Mname, Takings

Resulting relation has Fno and Fname filled in for each 'row

Key attributes are Fno,Cno

Part a Answer pointers

There was much confusion about how to identify a repeating group – many candidates thinking that Filmno and Fname were the repeating group

(b) Normalise the data, showing functional dependencies and how you progress from 1NF through 2NF to a set of 3NF relations.
At each stage show the primary key and any foreign keys of each relation and state assumptions that you make about any of the relationships between the columns of data.

(10 marks)

1NF table structure

Fno, Fname, Cno, Cname, CLocation, Mno, Mname, Takings

1NF to 2NF

Dependencies in 1NF table are

Fno -> Fname

Cno -> Cname, CLocation, Mno, Mname

Fno, Cno -> Takings

Remove part-key dependencies

Only column fully dependent on the key is Takings. The rest need to be removed, resulting in three tables

Film : Fno, Fname

Cinema : Cno, Cname, CLocation, Mno, Mname

Showing : Fno, Cno, Takings

2NF to 3NF

Remove transitive dependencies

Mname is transitively dependent on Cno via Mno

Cno -> Mno -> Mname

Remove the transitively dependent attribute, Mname.

Film : <u>Fno</u>, Fname

Cinema : <u>Cno</u>, Cname, CLocation, Mno

Showing : <u>Fno</u>, <u>Cno</u>, Takings

Manager : <u>Mno</u>, Mname

Part b Answer pointers

Many candidates were able to spot the simpler structures such as Film, but very few placed 'Takings' in a table with the combined key of Cno and Fno

(c) Foreign keys are normally designed to protect data in tables that are in a parent-child relationship. If a column in a table is declared to be a foreign key, what integrity checks will be placed on data in the two tables involved in the foreign key definition?

(4 marks)

Values in the foreign key column(s) must exist in the corresponding columns of the parents primary or unique key that is being referenced.

NULL values are an exception to this rule.

Key values of rows in the parent table that have are dependent foreign key values cannot be updated or deleted.

Part c Answer pointers

Disappointing to see that nearly all candidates did not seem to be aware of the constraints placed on changes to parent key values that have dependent child records. Few appreciated that foreign keys can be set to NULL.

(d) Many DBMS allow foreign keys to be defined with the CASCADE DELETE option. Explain the effect of this option on the two tables involved in the foreign key definition.

(3 marks)

This will allow rows in the parent table to be deleted even if there are dependent child records with matching foreign key values.

When the parent rows are deleted, all the dependent rows in the child table are also deleted as part of the transaction.

The delete action can be cascaded to second level dependents if ON DELETE CASCADE has been specified on the first level dependent table.

Part d Answer pointers

Relatively few candidates were able to explain this behaviour precisely. The cascading effect was very rarely mentioned

(e) Explain why it is customary to base primary keys on numeric (NUMBER) columns rather than character (VARCHAR) columns.

(4 marks)

VARCHAR columns tend to convey meaning, and hence are likely to be changed (update).

Cannot assign simple sequence numbers to VARCHARs to maintain uniqueness of primary key values. VARCHARs tend to be large and uncompressed, which means a rather bulky and less efficient index may need to be built to support the key. Numerics tend to small (possibly compressed) and system generated.

Part e Answer pointers

Candidates were aware that numeric columns tend to be unique, but few stated the sequencing and index/performance benefits. Very little was mentioned about the reduced likelihood of the need to update numeric keys

Question A2

This was a popular question – 85% of candidates attempted it.

A2. a) *Explain why* security of data in a database is becoming an increasing concern

[6 marks]

Increased level of attacks from both inside and outside. More sophisticated password crackers and database rootkits available. Reluctance of DBAs to patch database software. Government regulations requiring use of security features and techniques such as SOX, HIPAA, PCI-DSS. Increased auditor activity. Disclosure and data breach laws – fines for non-disclosure. Suspected Advanced Persistent Threats from alien governments. Extra security results in slower systems that may be less easy to use.

Part a Answer pointers

New environments like the internet and the problems they bring were well covered, but little was mentioned about government regulations and the problems of introducing security checks.

(b) Discuss techniques and precautions that can be used when

(i) allowing users to connect to the database.

(5 marks)

Complexity of passwords, length of passwords, secure storage and management of passwords. Identification of web-based users. – tracking through connection pools. Bio methods

(ii) authorising users to perform certain actions within the database

(5 marks)

Use of privileges and roles. Password protected roles. Role hierarchies. Invokers' and definers' rights. 'DBA' privileges. Application centric security vs database centric security. Row level security (fine-grained access).

(iii) Explain the effect of the following SQL statements when executed by a user called amy. Employees is a table owned by a user called amy

```
GRANT CREATE VIEW TO joe, fred;

REVOKE DELETE ON employees FROM joe;
```

(4 marks)

Joe and Fred can each now create views in their schemas.

Joe will no longer be able to delete data from any column in the employees table owned by amy. This will happen immediately.

Part b Answer pointers

Many candidates mentioned usernames and passwords, but few went on to discuss password management and other authentication technologies

Roles rarely received a mention. Many candidates did not seem to be aware of row-level security.

A vast array of inventive answers were submitted for part (iii) that showed a lack of understanding of privileges and database objects.

(c) To protect database data, backups of the database are taken. Discuss techniques and measures that can be taken in order to ensure the security and effectiveness of database backups.

(5 marks)

Encrypt data in the backup. Store in safe place away from database. Take multiple copies. Test and validate the backups. Perform trial recoveries. Password protect access to the backup. Use as part of a disaster recovery scenario. (construct hot-standby database).

Part c Answer pointers

This was answered better by the majority of candidates. Many were able to discuss the management of backups in some detail

Question A3

This was a reasonably popular question – 56% of candidates attempted it

A3. (a) In a distributed database, tables are often fragmented and/or replicated.

(i) Describe horizontal and vertical fragmentation of tables using examples where necessary and state reasons why it might be done. In each case show how the fragments may be recombined to return the entire set of data in the table.

[4 marks]

Horizontal slices the table on a row by row basis (often based on values of data in a column). Fragments must be recombined using `UNION` and uniqueness across fragments may be a problem

Vertical splits the table on the basis of columns. Fragments must be recombined using joins and primary keys must be reflected and maintained across the fragments

Part a(i) Answer pointers

Horizontal and vertical fragmentation were well understood but the operations required to recombine the fragments were not.

(ii) Describe replication of tables and discuss why this may be desirable and the problems it might cause.

[5 marks]

Performed for availability and performance reasons. Difficult to keep in sync and suffers from locking issues due to all replicas requiring locking. Also suffers from site failure issues when co-ordinating transactions

Part a(ii) Answer pointers

This sub-section was generally well answered with some good discussions

(iii) Describe the difference in behaviour when updating tables involved in

(1) synchronous replication

(2) asynchronous replication

[4 marks]

Synchronous requires two-phase commit and all sites to be available during the transaction. Could impact local operations. Asynchronous results in out of date replicas

Part a(iii) Answer pointers

Few candidates could explain the difference and the effects of these two types of synchronisation

(b) A main responsibility of the Database Administrator (DBA) is to ensure the availability of the database. To recover from failures that affect the database, the DBA uses a range of features such as after images, before images, transaction logs, checkpoints and backups within techniques such as rollforward, rollback and restore

Show how recovery may be achieved from the following failures.

(i) An update transaction is almost complete but fails to access the final table it needs to complete its changes

[3 marks]

The DBMS will discover the problem and apply appropriate before images from the undo area to roll back the effects of the transaction, thereby returning the database to a consistent state.

Part b(i) Answer pointers

The vast majority of candidates identified this as rollback – but few went on to talk about before images and consistency

(ii)A network connection fails resulting in the sudden crash of a database session conducted by a remote user.

[4 marks]

The DBMS will discover the problem and apply appropriate before images from the undo area to roll back the effects of the latest transaction to be submitted by the session, thereby returning the database to a consistent state. All of the session's usage of caches and locks will be cleaned out from memory.

Part b(ii) Answer pointers

Again, rollback was mentioned but nothing was said about the need to rollback only the current transaction being executed as of the time of the crash. Checkpoints were often mentioned – which usually gained credit

(iii) A non-mirrored disk belonging to the database overheats and all of its data is permanently lost.

[5 marks]

A previous backup will be restored on a valid disk and after images applied to roll the data forward. Then before images will be applied to roll back the effects of any transactions that were active as of the time of the crash.

Part b(iii) Answer pointers

Nearly all candidates highlighted the need to restore from backup, but it was disappointing to see only a few discussing the recovery option using after images, followed by as rollback of the changes due to in-flight transactions as of the time of the crash

Question B4

a) Define the term 'Relation' – in the context of the Relational model of data.

[3 marks]

Answer Pointers:

Relation: a table consisting of a set of rows (or tuple)s that have the same attributes. A tuple usually represents an object and information about that object. Data referenced by an attribute are in the same domain and conform to the same constraints.

Examiner's Comments:
In most answers, relation was confused with relationship (which is a link between relations).

b) Describe the function of each of the following :-
Views
Stored Procedures

[6 marks]

Answer Pointers:

Views provide abstraction of relations; they do not physically exist but are generated by a SQL query when the View is referenced. They can be used to: (i) reduce the complexity of the database for users by presenting only the data they need to see, (ii) provide a level of customisation by making the database appear differently to different users, (iii) provide a level of security by excluding data that some users should not see.

Stored procedures are subroutines (set of instructions written in SQL and other programming language) available to applications that access a database. As the name implies, a stored procedure is actually stored in the database. Typical uses for stored procedures include data validation or access control. But they can also be used to implement any business logic.

Examiner's Comments:
Although most students had a good attempt at defining the function of a view, very few seem to grasp the concept of stored procedure. This may be due to a lack of practical exposure to a DBMS product.

c) Write CREATE TABLE statements for a relational model derived from the Form given in Fig A1 (Appendix A). Include a small selection of attributes with data types and all foreign key constraints.

[8 marks]

Answer Pointers:

The answer should include more than one table, and should include relationships of some form between site, excavation and find. Each table should contain a selection of attributes but should include foreign keys to other tables. A possible answer would be as follows:

```
CREATE TABLE SITE(
siteNumber varchar(25) Primary key,
siteDescription varchar(250)…);

CREATE TABLE EXCAVATION(
ExcavationNumber varchar(5) Primary Key,
SiteNumber varchar(25) references Site(SiteNumber),
ChiefArchaeologist varchar(30)…);

CREATE TABLE  FINDS(
FindNumber integer,
FindType varchar(20),
ExcavationNumber references EXCAVATION(ExcavationNumber)…);
```

Examiner's Comments:
Most students made a satisfactory attempt at this question, although many produced only one table containing all the attributes. Some students may not have realised that "LIST OF FINDS" is part of Figure A1, given the line drawn between sites and finds.

d) Using the TABLES in part c) above, create an example View and an example Stored Procedure and show how they could implement a processing requirement described in the Case Study.

[8 marks]

Answer Pointers:

Full marks are given to a well-formed CREATE VIEW statement. The following is an example of a view that shows the find and excavation numbers related to chief archaeologist T.JOHNSON:

```
CREATE VIEW VWFINDEXCAVATED AS (
SELECT f.FindNumber, f.ExcavationNumber
FROM EXCAVATION e, FIND f
WHERE  e.ExcavationNumber = f. ExcavationNumber
AND e.ChiefArchaeologist = 'T.JOHNSON');
```

Full marks are given to a CREATE PROCEDURE statement that includes a reasonable set of statements including some SQL. The following is an example of a procedure that shows the site description corresponding to a given chief archaeologist. Note that the syntax used here is of PL/SQL, but it doesn't need to be:

```
CREATE OR REPLACE PROCEDURE ArchSite (IN chiefArch varchar(30))
AS
siteDesc SITE.siteDescription%TYPE;
```

```
BEGIN
SELECT siteDescription INTO siteDesc
FROM EXCAVATION e, SITE s
WHERE  e.SiteNumber = s. SiteNumber
AND e.ChiefArchaeologist = chiefArch);
Dbms_output.put_line('Site: '||siteDesc);
/
BEGIN
ArchSite('T.JOHNSON');
END;
```

Examiner's Comments:
It is important that students practice with the CREATE VIEW statement. Some students produced a table (view) filled with data instead of the actual statement that produces the view.

Almost all students failed to produce a reasonable procedure. Exposure to a commercial DBMS is vital in order to practice with writing such procedures.

Question B5

a) Explain with aid of examples what is meant by logical and physical data independence in a database system.

[10 marks]

Answer Pointers:
Logical data independence: The ability to change the logical (conceptual) schema without changing the External schema (User View) is called logical data independence. For example, the addition or removal of new entities, attributes, or relationships to the conceptual schema should be possible without having to change existing external schemas or having to rewrite existing application programs.

Physical data independence: The ability to change the physical schema without changing the logical schema is called physical data independence. For example, a change to the internal schema, such as using different file organisation or storage structures, storage devices, or indexing strategy, should be possible without having to change the conceptual schema.

Examiner's Comments:
Students should note that the drawing of the three-level architecture diagram does not count as an example of data independence.

b) Realisation of Logical and physical data independence may be difficult to achieve in practice in commercial DBMS products. Explain why this might be the case.

[5 marks]

Answer Pointers:

Logical data independence may be difficult to achieve because most databases rely on strong ties between the user view of the data and the actual structure of the underlying tables.

Physical data independence is compromised by having to tune physical database to represent more efficient joins for example. This leads to databases that are logically well normalised yet physically denormalised.

Examiner's Comments:
This question was very challenging to almost all students, with a limited few having made a satisfactory attempt. Students should develop an understanding of the limitations of the three-level architecture in real-life situations.

c) Most DBMS products use a data dictionary (also called the system catalogue) to hold meta- data. What is meta-data? Give examples of the type of data held in the data dictionary and explain what it is used for.

[10 marks]

Answer Pointers:

Meta data means data about data in other word in the context of data and database, it includes information such as the dates a database object (eg a table) was created, modified, or accessed, the author's name, the computer it was created on…
Catalog = Data Dictionary

Data Dictionary views present metadata in a format that is independent of any catalog table implementation, therefore catalog views are not affected by changes in the underlying catalog tables.

Important Views are provided that allows the internal database engine to use meta data such as sys_objects_catalog (or any other sensible example).

The number of columns in a table or view, together with the name, data type, scale, and precision of each column. The constraints that are defined on a table. The indexes and keys that are defined for a table

For example, the following query uses the sys.objects catalog view to return all database objects that have been modified in the last 10 days.

```
SELECT
  name AS object_name,
  SCHEMA_NAME(schema_id) AS schema_name,
  type_desc, create_date, modify_date
FROM sys.objects
WHERE modify_date > GETDATE() - 10
ORDER BY modify_date;
```

Examiner's Comments:
Most students failed to produce an example of a system view and a query that involves it, which may reflect the lack of practical exposure to a commercial DBMS.

Question B6

Examiner's Comments

This was the least popular question on the paper suggesting candidates struggle with applying data modelling techniques from a given scenario. There were 'easy' marks available as 10 marks were given for simply identifying

entity types and relationships between them. The model is quite simple and given the data in the appendix should return a data model with few assumptions. 10 marks were awarded for diagramming already identified entities and relationships with correct derived degree of each relationship. Generally many candidates still struggled with this question judged by the average mark which was below a pass and the fourth worse average mark across the paper. Worst performance was in the last part where a lack of understanding of participation and cardinality was an obvious weakness amongst most candidates.

Refer to the MSIS case study (Appendix A). Design an ER data model for the MSIS. The text highlighted (in bold font) is provided to help you identify entity types.

You are marked on the accuracy and expressiveness of your design and marked according to the following guidelines :

a Entity Types (5 marks)

b Relationship Types/name (5 marks)

c ER diagram (10 marks)

d A written description of ONE 1 to Many relationship from your ER diagram including an explanation of the participation constraints and modelling conventions that were used (5 marks)

State any assumptions you made – these must not contradict the discourse.

Answer pointer parts a-d

Entity types 1 mark each Max 5 – if at least first 3 listed below, +2 others correctly identified: -

FINDS, The main entity that defines a table of excavated items

SITE, The defined unique area that an excavation occured

EXCAVATION, The event of carrying out a excavation over a defined duration. Subsequent excavations could be undertaken

EXHIBITION The event that occurs to display certain FINDS presented over a defined duration

PERIOD OF OCCUPATION A reference list to allocate a SITE to a particular that a site was occupied – maybe many periods over which the site was occupied.

MEDIA SOURCE: As defined the type of multi-media that records the information about a site.

COLLECTION: Categorisation of FINDS and how they are organised ie mixed, individual and whole.

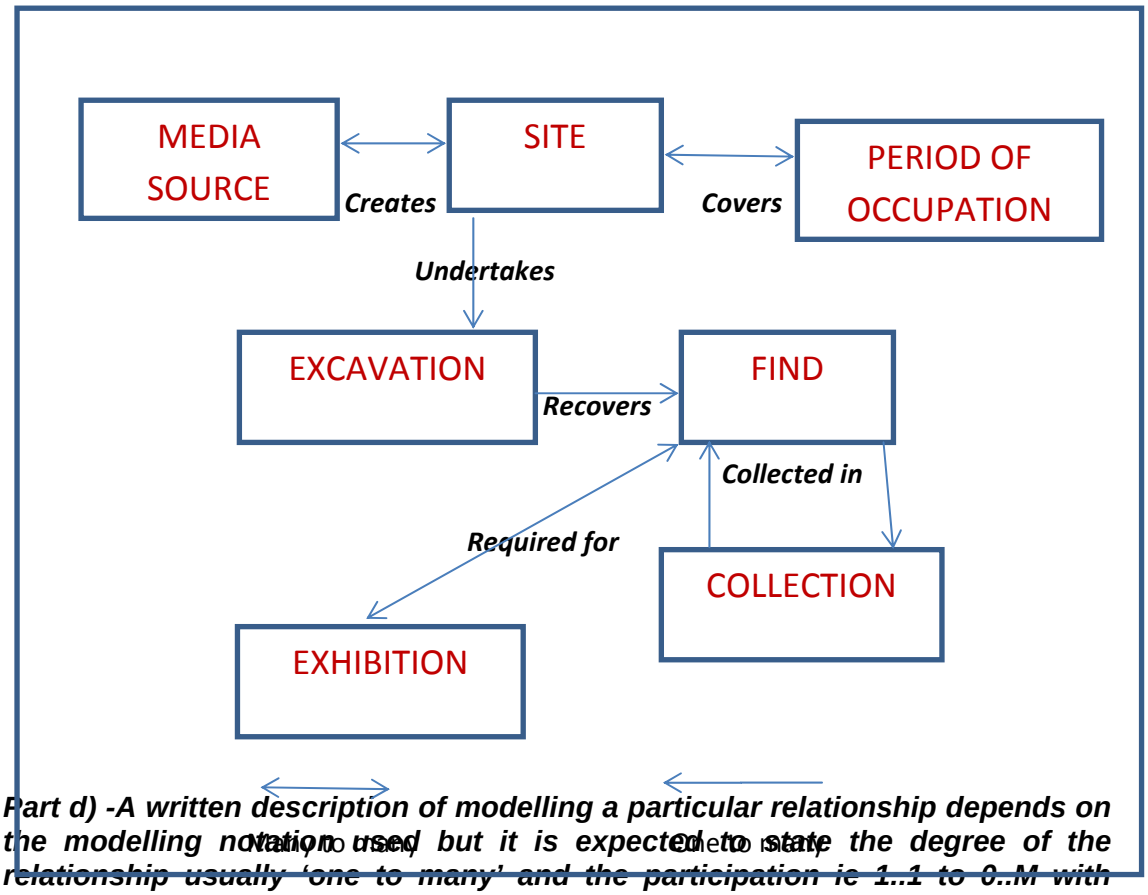
Relationship Types:-

These are named types to define relationships between entity types and can be either 1 to 1; 1 to many or many to many. For example check the data sources in the appendix and it is possible to derive that excavations occur at a particular site on different dates so the relationship type is excavate the degree is 1:many between site and excavation. Similarly there are many Finds excavated during a particular excavation. A Find may be exhibited at many different exhibitions on different dates and an exhibition exhibits many Finds.

Other relationship types are named as verbs usually HAS ; COVERS and are shown on the ERD.

The main modelling decision is to link FINDS with EXCAVATION rather than FINDS with SITE as there may be many EXCAVATIONS at a SITE and we need know which FIND was recovered at which EXCAVATION.

Please note at this level only simple explanation of the relationship between FINDS and COLLECTIONS which is 1:1 and has the name/type of 'collected in'. Participation constraints are expressed in the detail part iv) therefore the following diagram should be used to summarise the overall entity and relationship model without participation constraints.



Part d) -A written description of modelling a particular relationship depends on the modelling notation used but it is expected to state the degree of the relationship usually 'one to many' and the participation ie 1..1 to 0..M with some rationale. So 1..1 to 0..M would apply to SITE and EXCAVATION for example a SITE can have none, one or many excavations and an EXCAVATION must apply to one and only one site. Any diagram to express this would also be useful using a notation such as UML class notation.

Appendix A Case Study: (MSIS)

1. FINDS

The **Museums Service Information System** (MSIS) maintains a catalogue of about 5000 **Finds** excavated following an archaeological **excavation (or 'dig')**.

Figure A1 represents an example of cataloguing 2 finds that were excavated at a particular **site**.

A Find is recorded as either:

- **Individual:** a catalogued **item** that is intact (eg a sword)
- **Whole: a collection** of catalogued items (eg bones, tile) that either partially or completely that originally formed a single catalogued item (eg a skeleton, mosaic).
- **Mixed** a collection of non-catalogued items that are kept together as a single catalogued item. (eg shard of pottery). The collection may be of the same type (eg mixed pottery) and of similar age.

Once Finds have been catalogued they are stored at the district museum. **Artefacts** such as photographs, drawings from the excavation accompany each Find, either digitised, on paper or both (see Figure A1).

2. SITES

A Site is a location of special archaeological interest. Only a small proportion of sites have been excavated. But a record of the site is important to protect it from development.

3. USERS

A Chief Archaeologist is responsible for all aspects of an excavation and the catalogue of Finds. A Museum Service Curator is responsible for the upkeep of Finds and exhibiting a number of selected Finds at **exhibitions** at a museum. Archaeologists who are members of the excavation team fill in the form (fig A1) and can upload the information with any digital media to the Museum Service web site. The web site supports on-line access to summary information in the catalogue of Finds to anyone to search/browse the catalogue and get information about exhibitions.

Fig A1: Standard Data Input Form:

[**SiteNumber**] eg 1290/HG/4

[**Site spatial coordinates** in degrees] LAT 55.8672 LONG 23.9811

[**Excavation Number**] 1 of 1

[**Excavation Duration**] 12/Oct/98-22/Oct-98

[**Chief Archaeologist**] T.JOHNSON

[**Description of Site**] Early bronze age burial site with re-occupation afterwards mainly for farming. Also site of battle - 1066.

[**Period of Site Occupation Period1**] 2nd Century BC

[**Period of Site Occupation Period2**] 1066

[**Period of Site Occupation Period3**] 12th-14th Century

[**Media Sources LISTED BELOW**

X **Scaled line drawings of the position and location of the site.**

X **Colour illustrations (hand drawn) visualising the site.**

X+ **Colour 35mm photographic slides of a site.** (bitmap image format)

Videos (cassette tape) containing presentations on site history.

X+ **Aerial Photographs taken over the site.** (original Jpeg images)

Microfiche slides of old documents associated with a Site.]

==========

[LIST OF FINDS EXCAVATED AT THIS SITE]

[**Number of catalogued types=**] 2

[**FindNumber**] 9701

[**Type**] Mixed

[**Date of Find**] 180-220BC

[**Description**] Probable grave remains mainly bones but no skeleton

[**Cross Referenced FindNo**] 9702

[**Cross Reference Sitenumber**]

[**Comments**] an item removed from the grave remains

[**Amount/Units**]: 200 items of bones of mixed origin

Continued

[**FindNumber**] 9702

[**Type**] Individual

[**Date of Find**] 180BC

[**Description**] 24ct Gold Tunic ring from a chieftain

[**Cross Referenced FindNo**] 826

[**Cross Reference Sitenumber**] HT98/96

[**Comments]** similar ring found at the above site

[**Amount**]