The University Of Sheffield.

Department Of Computer Science.

# Handling Data

Mike Stannett
Department of Computer Science
University of Sheffield

Handsworth Grange Community Sports College | Westfield School | All Saints | sero | MADE IN SHEFFIELD

## Aims for today

- To explain how data is used and manipulated by online applications, and the legal framework for doing so
- To explain how data can be encrypted for security purposes, and techniques for cracking codes
- Working in groups:
  - Encrypt a message and store it online
  - Decrypt another group's message (if you can!)

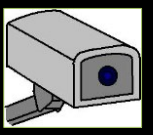
2

Before we start … what can and can't you do with data?

## RIGHTS AND RESPONSIBILITIES

## What do you get if you combine...

camera



*Microsoft Visio*

photographic database

*http://www.3dface.org/media/images.html*

facial recognition software

## Dataveillance

1. Take photo

2. Extract features

4. Real-time surveillance

3. Identify

*http://www.3dface.org/media/images.html*

## Ethical issues

- What effect would the existence of such a system have on society?
- Is it OK for such a system to be used by the police? What about other authorities?
- Is it possible to stop criminals getting access to such a system if it exists? What about terrorists?
- Is it ethically acceptable to build such a system?
- Is it ethically acceptable to do research that would make such a system possible?

## Concerns over privacy

- Governments and businesses store large amounts of personal information
- Not obvious that organisations really need the information they hold
- Databases can be searched rapidly to provide profiles of individuals (shopping preferences, DNA, favourite travel routes, banking habits, ...)
- This information can be sold (advertisers) or stolen (criminals, identity theft)
- Disrupts balance between state and individual

Following an EU Directive, the Data Protection Act 1998 came into force in March 2000.

Data Protection Act 1998

1998 CHAPTER 29

ARRANGEMENT OF SECTIONS

Go to Preamble

PART I

PRELIMINARY

1. Basic interpretative provisions.
2. Sensitive personal data.
3. The special purposes.
4. The data protection principles.
5. Application of Act.
6. The Commissioner and the Tribunal.

http://www.opsi.gov.uk/acts/acts1998/plain/ukpga_19980029_en_1

## What the Act does

- Sets out 8 principles, and aims "to protect the fundamental rights and freedoms of natural persons, in particular their right to privacy with respect to the processing of personal data".
  - gives legal rights to individuals (e.g. to have incorrect data removed or corrected)
  - appoints an Information Commissioner, who
    - maintains a list of who holds personal data
    - serves notice on those who contravene the Act
    - ensures that requests from individuals to persons who hold data about them are honoured

## DPA: First Principle

- Personal data shall be processed fairly and lawfully, and shall not be processed unless at least one of the following is satisfied:
  - the data subject has consented
  - the processing is necessary for the performance of a contract with the data subject
  - the controller has a legal obligation to process the data
  - the processing is necessary for legitimate interests pursued by the data controller
- Consent need not be explicit (failure to tick a box on a form), unless the data is "sensitive" (racial origin, political opinion, sex life, ...)

## DPA: Second Principle

- Personal data shall be obtained only for one or more specified and lawful purposes, and shall not be further processed in any manner incompatible with those purposes.

*Example of breach*: Your gas supplier has your permission to send you bills and material relating to its own gas-supply services, but starts including other adverts as well.

## DPA Third & Fourth Principle

- Personal data shall be adequate, relevant and not excessive in relation to the purposes for which it is processed.

Personal data shall be accurate and, where necessary, kept up to date.

"It is not acceptable to hold information on the basis that it might possibly be useful in the future without a view of how it will be used."

## DPA Fifth & Sixth Principle

- Personal data processed for any purposes shall not be kept longer than is necessary for those purposes.

Example: Universities set a time limit, after which old exam scripts must be destroyed.

Personal data shall be processed in accordance with the rights of data subjects under the Act.

## DPA Seventh Principle

- Appropriate technical and organisational measures shall be taken against unauthorised or unlawful processing of personal data and against accidental loss and damage
- A data controller must choose a data processor who provides sufficient guarantees in this respect (and there must be a written contract ensuring this).

## DPA Eighth Principle

- Personal data shall not be transferred to a territory outside the EU unless that territory ensures an adequate level of protection for the rights and freedoms of data subjects
  - Exception: if consent has been obtained

## Your rights under the DPA

- Entitled to know if data is held about you
  - If so, entitled to know what this data is
- Entitled to know the logic used by any fully automated decisions making process
  - can insist this isn't done
- You can require them not to process data that's likely to cause damage or distress
  - may be entitled to compensation
- Can insist data isn't used for direct marketing
- Can insist that errors are amended

## Important exemptions (there are others)

Personal data is exempt from the Act if processed

- for purposes of national security
- for journalistic, literary or artistic purposes in the public interest

**Exemptions relating to Examinations**

Exam scripts are exempt from subject access

Access to exam grades can be delayed

## DPA Summary

Data Protection Act 1998

1998 CHAPTER 29

ARRANGEMENT OF SECTIONS

Go to Preamble

PART I

PRELIMINARY

1. Basic interpretative provisions.
2. Sensitive personal data.
3. The special purposes.
4. The data protection principles.
5. Application of Act.
6. The Commissioner and the Tribunal.

People have rights

You have responsibilities

Note: The Act applies to manual filing systems as well as electronic ones.

Other Acts deal with the interception of communications data, surveillance techniques, and disclosure of encrypted data. Example: ISPs have to maintain interception capability.

Server-side languages are used to manipulate web pages before they are transmitted to the browser.

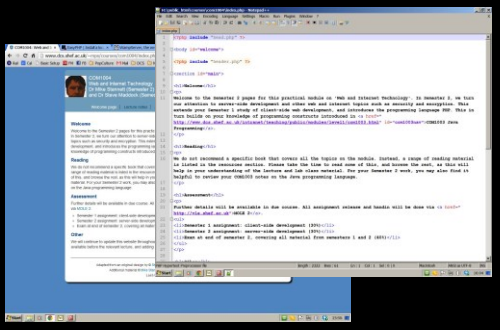
## SERVER-SIDE LANGUAGES

19

## Teach Yourself W3C Schools: PHP Tutorial


20

## PHP in action



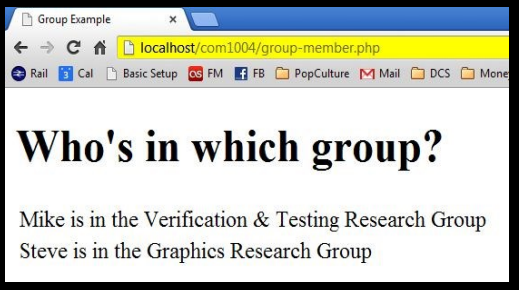
21

## Why does this code...

```
<!DOCTYPE html>
<html lang="en">
<head>
  <meta charset="utf-8" />
  <title>Group Example</title>
</head>
<body>

<h1>Who's in which group?</h1>
<?php
  define("VT", "Verification &amp; Testing");
  $group = array( "Mike" => VT, "Steve" => "Graphics" );
?>
<table>
<?php
  foreach ($group as $key => $value) {
    echo "<tr><td>$key is in the $value Research Group</td></tr>";
  }
?>
</table>
</body></html>
```

22

## ... produce this web page?




23

## What is PHP?

- PHP = "PHP: Hypertext Preprocessor"
- Server-side scripting language
- Can be used as a stand-alone programming language
- Often used with MySQL, a database-handling language
- Useful for simplifying repetitive tasks
- Useful for building interactive websites

24

## Client-side vs server-side

- Client-side execution
  - JavaScript is transferred to your machine and then executed by your web client.
  - Any passwords in the code are clearly visible to the user, as well as to snoopers watching the net.
- Server-side execution
  - PHP scripts are executed on the server, and *only the results* are sent to your machine.
  - Only the server sees your private data; it is never transmitted over the net.

25

## Using PHP with HTML

- Web pages are written in a language called HTML
- To insert PHP commands into your web page, you use the following tags

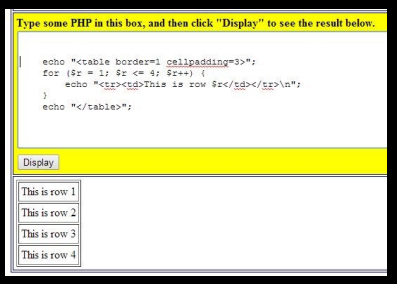
```
<?php
    (this is where you insert your PHP code)
?>
```

- As usual with HTML pages, the tags have to match.

26

## Using PHP with HTML

You can insert any valid PHP code between the tags




27

## Adding text to your HTML file

- To add something into the web page, use
  - echo "whatever";
- You can also use
  - print("whatever");

```
$header =
"<html><head><title>Example</title></head><body>";
$footer = "</body></html>";

$page = $header . "Some page contents" . $footer;
echo $page;
```

28

## Quoting conventions

- Strings can be enclosed in single or double quotes.

```
$h = "hello";
$t = 'there';
```

- If you include a variable inside double-quotes, it'll be expanded

```
$h = "hello";
$t = 'there';
echo "$h $t";  // prints hello there
echo '$h $t';  // prints $h $t
```

29

## Joining strings together

- You can use quotes to join values together within strings

```
$h = "hi";
$f = 5;

echo "$h $f";  // prints hi 5
```

- You can also use the concatenation operator for strings

```
$h = "hello";
$t = 'there';

echo $h . " " . $t;  // prints hi
there
```

30

## .. and so on …

- PHP is similar to many other programming languages, and just as powerful
- It has variables, constants, arrays, built-in functions, etc., and the way you write programs is much the same as with other languages
- PHP has a lot of built-in functions that are especially designed to make data handling easy

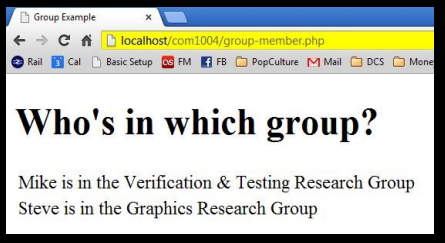
31

## Example: PHP inside HTML

```
<!DOCTYPE html>
<html lang="en">
<head>
  <meta charset="utf-8" />
  <title>Group Example</title>
</head>
<body>
  <h1>Who's in which group?</h1>
  <?php
    define("VT", "Verification &amp; Testing");
    $group = array( "Mike" => VT, "Steve" => "Graphics" );
  ?>
  <table>
  <?php
    foreach ($group as $key => $value) {
      echo "<tr><td>$key is in the $value Research Group</td></tr>";
    }
  ?>
  </table>
</body></html>
```

32

## Example




33

## ADDING DYNAMIC CONTENT

34

## What is dynamic content?

- Standard web pages are the same every time you visit them.
- Dynamic content can change to reflect
  - the time of day
  - the IP address of the person viewing the page
  - personal viewing preferences
  - the web browser being used
  - information saved in a shopping basket during earlier visits
  - …

35

## Adding a date stamp to your page

- Example: News stories often change quite quickly
  - Users need to know whether information is up-to-date
  - Adding a date stamp lets them know when the page was last updated.

```
$mtime =
      filemtime($_SERVER['SCRIPT_FILENAME']);
print date("D d M Y", $mtime);
```

36

## Adding the current time to your page

- When you order something on an e-commerce site, the site often needs to make a note of when you confirmed the order.

```
echo "<p>Order processed at ";
echo date('H:i, jS F Y');
echo "</p>";
```

The `date()` function is used to print both dates and times.

37

## Why can we access dynamic information?

- Since PHP is executed on the server, it can access
  - values of local environment variables
  - the time recorded on the server's clock
  - the local file system, and more ...
- Since PHP is part of a web page requested by the client, it can access
  - the viewer's IP address
  - which browser the user is using to view the page
  - and more ...

38

## phpinfo

- The **`phpinfo`** function tells you
  - version information for PHP
  - about the web server
  - what browser this page is being viewed on
  - what this page's file path is on the server
  - the URL of the page that requested this page
  - the HTTP headers used to request this page
  - and lots, lots more...

39

## Some of the information provided by phpinfo

```
User-Agent Mozilla/5.0 (Windows; U; Windows NT 5.1; en-GB;
  rv:1.9.2) Gecko/20100115 Firefox/3.6 (.NET CLR
  3.5.30729)
Referer http://localhost/com1004/code-
  tester.php?code=phpinfo%2863%29%3B%0D%0A
DOCUMENT_ROOT C:/Documents and Settings/Mike/My
  Documents/WWW
PHP Version 5.3.0
Calendar support enabled
QUERY_STRING code=phpinfo%28%29%3B%0D%0A
FTP support enabled
Multibyte (japanese) regex support enabled
Session Support enabled
```

40

## Using HTML forms

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN">
<html>
  <head>
    <meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
    <title>PHP Example</title>
  </head>
  <body>
    <form name="myform"
          action="javascript:alert('hi there, ' + document.myform.who.value);">
      <table border="1" cellpadding="4">
        <tr><td align="right">Your name:</td>
            <td><input name="who" type="text" maxlength="40" size="10"></td>
        </tr>
        <tr><td colspan="2"><input type="submit" value=" Click me! "></td></tr>
      </table>
    </form>
  </body>
</html>
```

41

## HTML forms

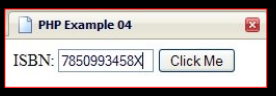
- Adding a form to a web page

```
<form
  name="myName"
  action="howToHandleTheData"
  method="howToSendTheData">

<!-- input boxes go here -->

</form>
```

42

## HTML forms

```
<form name="order" action="process.php" method="get">
ISBN:
  <input type="text" name="isbn" maxlength="11" size="11" />
  <input type="submit" value=" Click Me " />
</form>
```




When the user clicks the button, a request is sent using this URL

```
process.php?isbn=7850993458X
```

43

## How PHP makes form data available

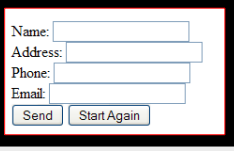
```
process.php?isbn=7850993458X
```

When this URL is called, the following array variable is automatically defined inside `process.php`

```
$_REQUEST["isbn"]  = "7850993458X";
```

44

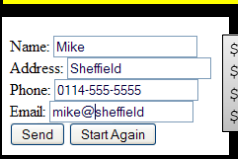
## Passing lots of data




```
<form action="process.php">
Name: <input name="name" type="text"><br />
Address: <input name="address" type="text"><br />
Phone: <input name="phone" type="text"><br />
Email: <input name="email" type="text"><br />
<input type="submit" value=" Send " />
<input type="reset" value=" Start Again " />
</form>
```

45

```
process.php?name=Mike&address=Sheffield&phone=0114-555-5555&email=mike%40sheffield
```




```
$_REQUEST["name"]
$_REQUEST["address"]
$_REQUEST["phone"]
$_REQUEST["email"]
```

```
<form action="process.php">
Name: <input name="name" type="text"><br />
Address: <input name="address" type="text"><br />
Phone: <input name="phone" type="text"><br />
Email: <input name="email" type="text"><br />
<input type="submit" value=" Send " />
<input type="reset" value=" Start Again " />
</form>
```

46

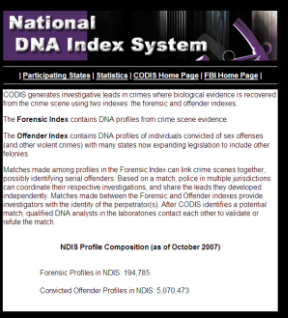
How data is stored for easy access

# DATABASES AND THEIR USES

## Database = ?

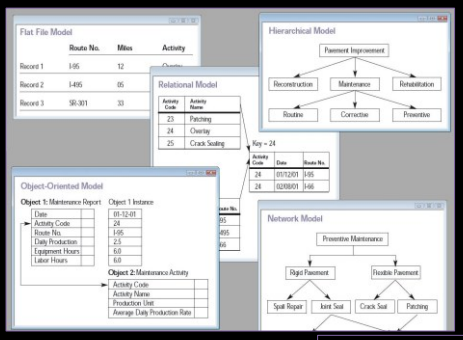
*Database*: a collection of records that
- are related to one another in some logical way
- can be regarded as a single collection
- provide data for one or more users

*Database classification*:
- type of content
  - bibliographic, DNA, …
- database model
  - relational
  - hierarchical
  - networked




CODIS: http://www.fbi.gov/hq/lab/codis/national.htm

## Database models




Marcel Douwe Dekker (Wikimedia Commons)
http://en.wikipedia.org/wiki/Database_model

## Flat files

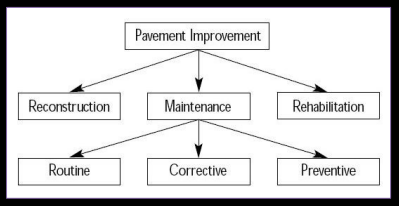
In the flat model, all data is stored in a single 2d array (each record is written to a new row in the table)
Example: using a text file to store a list of orders.

Flat File Model

| | Route No. | Miles | Activity |
|---|---|---|---|
| Record 1 | I-95 | 12 | Overlay |
| Record 2 | I-495 | 05 | Patching |
| Record 3 | SR-301 | 33 | Crack seal |

Wgabrie (Wikimedia Commons)
http://en.wikipedia.org/wiki/Database_model

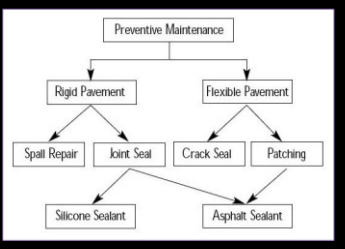
## Hierarchical model




This is the standard 'organisational chart' style of storing data. Data is stored in a tree-like structure.

U.S. Department of Transportation
http://en.wikipedia.org/wiki/Database_model

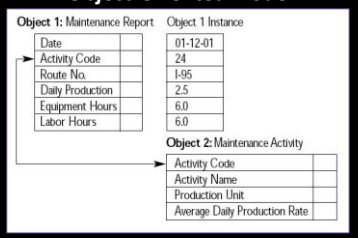
## Network model




In this model there may be multiple routes one data entry to another.

U.S. Department of Transportation
http://en.wikipedia.org/wiki/Database_model

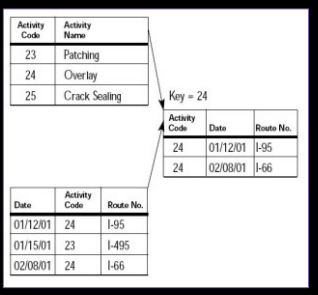
## Object-Oriented model




Data is stored in a way that reflects the class structure of the (object-oriented) program that uses it.

U.S. Department of Transportation
http://en.wikipedia.org/wiki/Database_model

## Relational model




Data is split across multiple tables.

Keys are used to link data in one table with data in another.

U.S. Department of Transportation
http://en.wikipedia.org/wiki/Database_model

Storing large data collections efficiently

# RELATIONAL DATABASES

## What they're for

Relational databases provide

- faster access to data than flat files
- can be queried to extract data satisfying specific conditions
- handle concurrent access - no need to worry about it
- provide random access to data
- have built-in privilege systems

Flat files are quicker to set up,
so can be better in some situations.

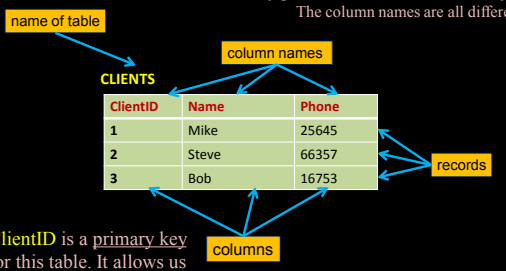
## Tables (aka Relations)

- In pure mathematics, a relation is a kind of "many-to-many" function, and can be written as a table. The underlying maths can be used to streamline data access.

| x | y |
|---|---|
| 0 | 2 |
| 1 | 3 |
| 2 | 3 |
| 2 | 16 |

## Table structure

The data in any given column are all of the same type.
The column names are all different.




ClientID is a primary key for this table. It allows us to identify records in the table.

CLIENTS(ClientID, Name, Phone)

## Tables are linked via keys

Primary keys are underlined, and foreign keys are shown in *italics*.

CLIENTS(ClientID, Name, Phone)

| ClientID | Name | Phone |
|---|---|---|
| 1 | Mike | 25645 |
| 2 | Steve | 66357 |
| 3 | Bob | 16753 |

ORDERS(OrderID, *ClientID*, Feedback)

| OrderID | ClientID | Feedback |
|---|---|---|
| 1034562 | 1 | Great product - five stars. |
| 5637265 | 3 | Please send more bricks. |
| 9864789 | 1 | Arrived late, and didn't work. |

ClientID is a foreign key in this table. It allows us to select rows in the other table.

## Following the links

CLIENTS

| ClientID | Name | Phone |
|---|---|---|
| 1 | Mike | 25645 |
| 2 | Steve | 66357 |
| 3 | Bob | 16753 |

ORDERS

| OrderID | ClientID | Feedback |
|---|---|---|
| 1034562 | 1 | Great product - five stars. |
| 5637265 | 3 | Hi - I'm a builder. |
| 9864789 | 1 | Arrived late, and didn't work. |

Suppose we need to contact the client who had trouble with order 9864789. What phone number should we ring? This is called a query.

## Following the links

CLIENTS

| ClientID | Name | Phone |
|---|---|---|
| 1 | Mike | 25645 |
| 2 | Steve | 66357 |
| 3 | Bob | 16753 |

ORDERS

| OrderID | *ClientID* | Feedback |
|---|---|---|
| 1034562 | 1 | Great product - five stars. |
| 5637265 | 3 | Hi - I'm a builder. |
| 9864789 | 1 | Arrived late, and didn't work. |

Queries are written in a language called SQL (Structured Query Language)

```
SELECT CLIENTS.Phone from CLIENTS, ORDERS
where CLIENTS.ClientID = ORDERS.ClientID
and ORDERS.OrderID = 9864789;
```

## Free Systems for Handling Data

- PHP
  - is a language commonly used for writing programs that modify web pages before they are delivered to browsers
- MySQL
  - is a commonly used free version of SQL
- Try it yourself
  - PHP: http://www.php.net/
  - MySQL: http://www.mysql.com/
  - WAMP (all in one): http://www.wampserver.com/en/

62

## DESIGNING YOUR DATABASE

63

## Choose your tables carefully

| isbn | author | title | cost | orderid | customerid | date | name | address |
|---|---|---|---|---|---|---|---|---|
| 089.. | Newton | Pr... | £20 | 67 | 2985 | 10th | Mike | sheffield |
| 089.. | Newton | Pr... | £20 | 68 | 546 | 17th | Steve | sheffield |

This structure uses 1 table, which wastes space.
Lots of information is repeated from one row to the next.

Avoid adding columns that will mostly be empty, e.g. a reviews column. Add a new table instead, and only create the relevant rows.

| *isbn* | customerid | review |
|---|---|---|
| 089.. | 2985 | Classic! |

## Choose your tables carefully

```
SELECT books.cost FROM customers, books, orders
WHERE books.author = "Newton"
AND books.isbn = orders.isbn
AND orders.customerid = customers.customerid
AND customers.name = "Mike";
```

books

| isbn | author | title | cost |
|---|---|---|---|
| 089.. | Newton | Pr... | £20 |

This structure uses 3 fairly compact tables. All of the cells in each row contain values.

orders

| orderid | *customerid* | *isbn* | date |
|---|---|---|---|
| 67 | 2985 | 089.. | 10th |
| 68 | 546 | 089.. | 17th |

customers

| *customerid* | name | address |
|---|---|---|
| 546 | Mike | sheffield |
| 2985 | Steve | sheffield |

## To create a new table in MySQL

```
CREATE TABLE customers
( customerid INT UNSIGNED NOT NULL
             AUTO_INCREMENT PRIMARY KEY,
  name CHAR(50) NOT NULL,
  address CHAR(100) NOT NULL
);
```

To create a table, you say "create table", then give the table's name, and then say what each column is called and what type of data it holds.

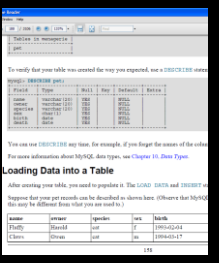
## Using MySQL inside PHP

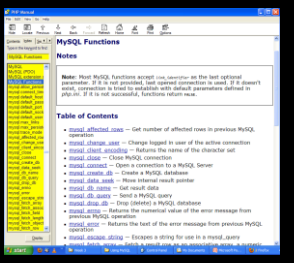
```
$db = new mysqli('host', 'user', 'pass');
if (mysqli_connect_errno()) {
   echo "Error: Couldn't connect";
   exit;
}
$db->select_db('dbname');
$query = "SELECT ..... ;";
$result = $db->query($query);
...
$result->free();

$db->close();
```

See the PHP and MySQL websites for downloadable manuals

## Further Reading

- "Chapter 3: Tutorial" in the MySQL manual
- "MySQL Functions" in the PHP manual




## Where to find the software

- PHP
  - Used to manipulate web pages before they're downloaded
  - www.php.net
- MySQL
  - A commonly used free version of SQL
  - www.mysql.com

WAMPs (all in one: Windows, Apache, MySQL & PHP)

EasyPHP: http://www.easyphp.org/
WampServer: http://www.wampserver.com/en/

69

## USING DATA WITH WEB PAGES

70

## Saving and retrieving data

- When a user submits a web form, how do we save the information to a database?
- How do we retrieve that data later?

| ID | NAME | EMAIL | MESSAGE | TIMESTAMP |
|---|---|---|---|---|
| 1 | Mike | mike@dcs | Hi there | 2014-03-03 10:31:04 |
| 2 | Steve | steve@dcs | Gotcha | 2014-03-03 10:33:07 |
| 3 | Mike | mike@dcs | Great - it works! | 2014-03-04 08:17:35 |

## From form to database

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN">
<html>
<head>
 <meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
 <title>MySQL - Exercise 1</title>
 <?php
 /* The following "info file" is not stored under "public_html", and
  * cannot be accessed online by specifying a URL. It defines the values of
  *   $myDbase   --- my database (only used for this example)
  *   $myPass    --- my password (only used for this example - NEVER share
  *                  your real password with anyone!)
  *   $myName    --- my username (only used for this example)
  *   $myHost    --- my hostname (only used for this example)
  */
 $info = "/mysecretpath/info.php";
 include $info; // this command inserts the contents of my info script
                // into THIS page, at this point in the file.
 ?>
</head>
<body>

<!--
  Create a form with data that needs to be stored or retreived. If the
  user clicks the "Retrieve" button, use JavaScript to cehck that they
  have actually entered an email address (we'll be retrieving all the
  messages associated with that address). If the user clicks the "Submit"
  button, make sure that both the email and the msg fields have been
  completed, since there won't be anything sensible to store otherwise.
-->
```

## Step 1. What data are involved?

- Design the form
  - Identify the information you want the user to provide
  - Create <input> tags accordingly
  - Here we want the user to provide Name, Email and Message data


## Step 1. What data are involved?

- Design a MySQL table to hold the data
  - Draw a picture of the required database table(s)
  - There should be a column for each input tag in the form that contain the relevant data
  - You may want to store extra data that isn't in the form (e.g. a timestamp and message ID)

| ID | NAME | EMAIL | MESSAGE | TIMESTAMP |
|---|---|---|---|---|
| 1 | Mike | mike@dcs | Hi there | 2014-03-03 10:31:04 |
| 2 | Steve | steve@dcs | Gotcha | 2014-03-03 10:33:07 |
| 3 | Mike | mike@dcs | Great - it works! | 2014-03-04 08:17:35 |

## Step 2. Make sure the table exists

- Does the table exist?
  - Don't assume it exists just because you created it earlier - it may have been deleted in the mean time…
  - If necessary, create the table

```
CREATE TABLE IF NOT EXISTS
  mytable (col1, col2, …);
```

## Step 3. Write code for each required activity

### Submit data

- If the user provided data, check it makes sense
- Create an INSERT query to add data to the table
- Submit the query to MySQL
- Check for success

### Retrieve data

- If the user provided data, check it makes sense
- Create a SELECT query to get data from the table
- Submit the query to MySQL
- Check for success

## Step 4. Keep the user informed

- Your request has been processed; you will hear from us within 7 days
- Sorry, we couldn't handle your request. Please try again later.

- The data you requested is not available. Please try again later.
- Here is the data you requested.

## Don't forget security!

- Malicious code injection. Consider this:

  "SELECT * FROM users WHERE
  name = '" + **$userName** + "';"

- If the user says their name is

  me'; DROP TABLE users;
  SELECT * FROM private WHERE '0' = '0

- This becomes

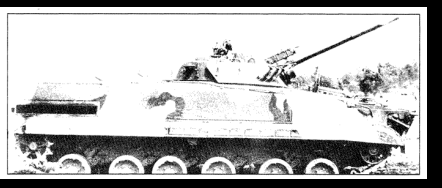
  SELECT * FROM users WHERE name = 'me';
  DROP TABLE users;
  SELECT * FROM private WHERE '0' = '0';

This deletes the **users** table and
lists the entire contents of the **private** table

# ONLINE SECURITY

79

## What are the threats?



- All businesses face threats
  - They can be systemic (bad planning, bad structure)
  - They can be internal (unhappy staff, lack of training)
  - They can be external (illegal attacks, competitors)

http://commons.wikimedia.org/wiki/File:Bmp-3_scanned_side_view.gif

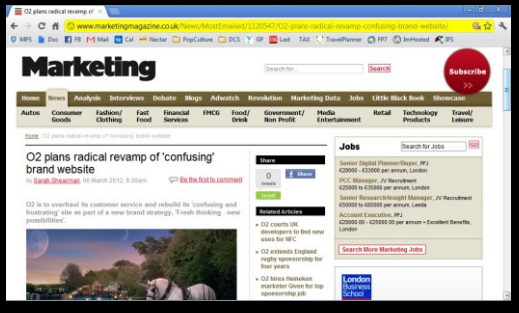
## Important business threats

- Hackers
- Not getting enough business
- Getting too much business
- Hardware failure
- Network and utility failure
- Reliance on other companies
- Competition
- Software errors
- Government policies and taxes
- Capacity limitations

## Important security threats

- Exposing confidential data
- Loss or destruction of data
- Modification of data
- Denial of Service (DoS, DDoS) attacks
- Errors in software
- Repudiation

## Confusing your visitors

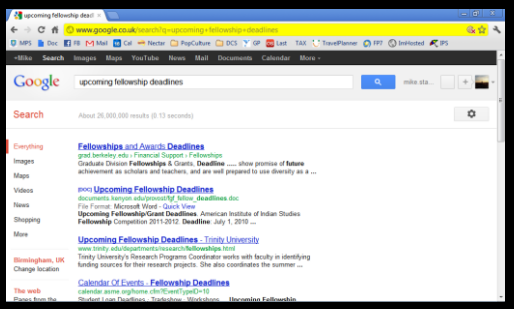



http://www.marketingmagazine.co.uk/News/MostEmailed/1120547/O2-plans-radical-revamp-confusing-brand-website/

## Giving standardised answers

Hi - You haven't actually answered my question AT ALL. I already knew that the existing card couldn't be de-registered (I even said so in my original message). The question is whether the NEW card has to be registered....

## Outdated information




http://www.google.co.uk/search?q=upcoming+fellowship+deadlines

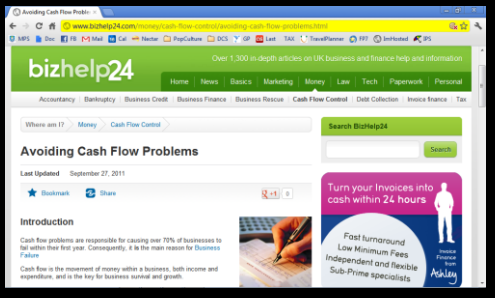
## Repudiation

Dear Amazon,

You have invoiced my credit card for a CD I never ordered, and which I have never received. Please return my money immediately.

How can Amazon prove you ordered and received the CD? Is it worth the cost of doing so?

## Succeeding too quickly

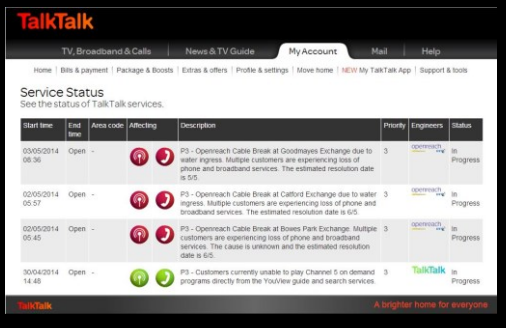



http://www.bizhelp24.com/money/cash-flow-control/avoiding-cash-flow-problems.html

## Remember to monitor and plan ahead

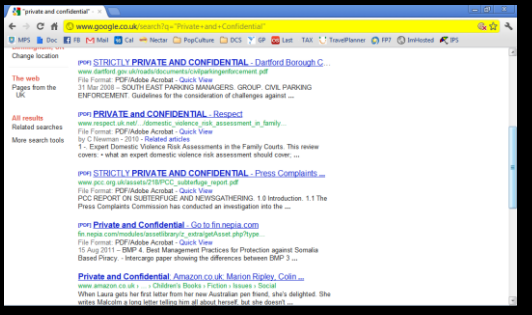
- How well are you doing?
  - Are you going to have cash-flow problems?
  - Where is future growth going to come from?
  - Who are you main customers?
  - Which customers are you missing out on?
- How can you improve?
  - Test to destruction?
  - Kanban / Just-in-Time distribution?

## Outages




http://www.talktalk.co.uk/servicestatus/

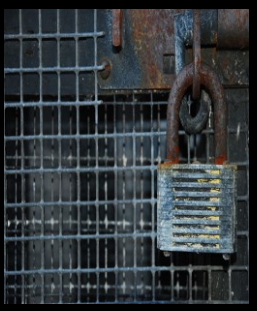
## Exposure of confidential information




http://www.google.co.uk/search?q="Private+and+Confidential"

## Denial of Service (DoS, DDoS)

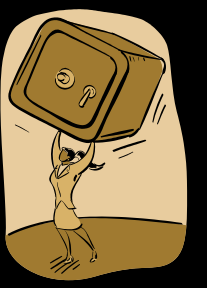
- Site may become inaccessible
- Access may slow down
- May affect wider network
- May affect wider region



## Companies need a security policy!

How will you cope with

- An earthquake
- DDoS attacks
- IT boss goes on holiday
- Fire
- Disk crash
- Broadband cable is cut
- Power cut
- Unhappy employees



Pretending to be something you're not

## GENERAL SCAMS

## "Nigerian 419"

Beloved in God

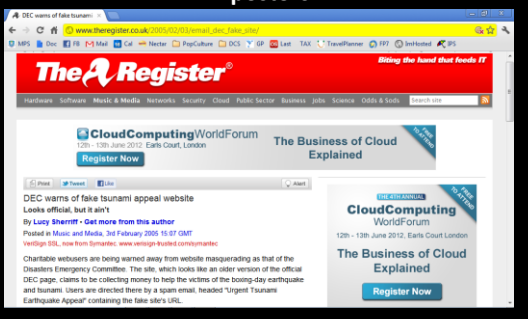
I am Rebecca, widow of the late MR WILLIAM SMITH Minister for Oil. I have urgently need transfer $12,345,674,201 to overseas for avoid death taxes. I urge your help, confident in God, and will pay fee 10%.

**BEWARE**: You cannot deduce anything from the way the email is written. It contains deliberate mistakes to make you think the sender is too naïve to bother about.

## Scam-baiting

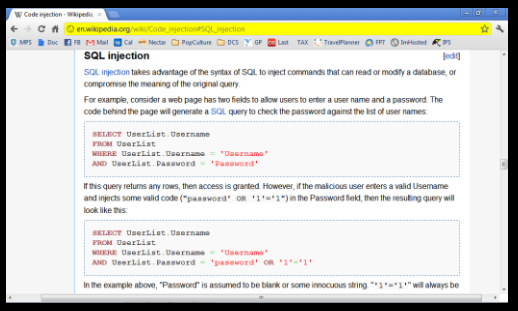

http://www.419eater.com/index.php

## Impostors



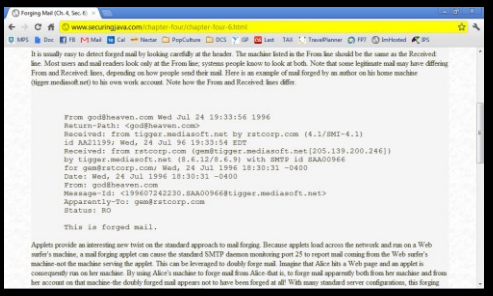

http://www.theregister.co.uk/2005/02/03/email_dec_fake_site/

## Malicious code injection




http://en.wikipedia.org/wiki/Code_injection#SQL_injection

## Forged email

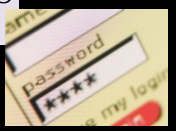



http://www.securingjava.com/chapter-four/chapter-four-6.html

## Phishing

- Forged emails can fool people into providing personal information

WE ARE INTRODUCING A NEW ONLINE SECURITY SYSTEM. CLICK HERE TO UPDATE YOUR PASSWORD.




## Executable attachments

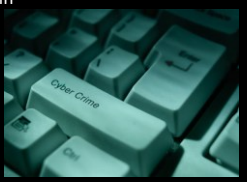
- Suppose you receive an attachment called "hello.jpg"
- What type of file is it?
- Is it safe to open it?


Windows lets you hide filename extensions for common file types, e.g. ".txt". The attached file could actually be called "hello.jpg.exe".

## COMPUTER AND INTERNET FORENSICS

## Computer crime

- Includes
  - Any crime that uses a computer in its commission
  - Hacking into computers
  - Downloading of illegal images
  - Accessing digital data without permission
- How do we prove who did what?



## Computer forensics

- Examine the suspect's computer and disks
  - Check existing files
  - Check emails and download logs
  - Examine .torrent files
- Reconstruct and check deleted files
  - Examine swap space
  - Look for file headers and familiar patterns
  - Magnetic properties of media can be exploited

## Internet forensics

- Internet traffic uses headers
- Some headers can't be forged easily

Received: from fake.gov
(mike.com [207.977.21.22])

## Real Example

- A student advertised on the web saying he was a stand-in lecturer, and needed urgent help finding the solution to an assignment, as the original lecturer hadn't left any notes
- The student placed several ads, using different names and email addresses, and offered to pay for a solution
- How did they identify the student?
- How long did it take?

Mathematics in action

## BASIC CRYPTOLOGY

106

## Basic cryptography

- History and relevance
  - Famous problems: Kryptos
- Basic Encryption and Decryption Techniques
  - Codes and ciphers
  - Caesar cipher, Vigenère cipher
  - Frequency analysis, cracking the Vigenère cipher
- Public-Key Systems: RSA, PGP

107

## Codes or ciphers?

"**CIPHER** denotes likewise certain secret characters disguised and varied, used in writing letters that contain some secret, not to be understood but by those between whom the cipher is agreed upon."
*Encyclopaedia Britannica*, 1st edition, 1771.

**CODE**:
"The blue moon rose early tonight"
==>
"The suspect left the building at dawn"

It's not the spelling that matters, but the meaning of the words. Codes may use a code book, explaining what phrase to use in different circumstances.

108

## Terminology

ENCRYPTION: Converting an intelligible message (plaintext) into something that can't be directly interpreted (ciphertext).

DECRYPTION: Converting a ciphertext back into the original plaintext.

CRYPTOSYSTEM: The underlying algorithms used to encrypt and decrypt messages.

KEY: Information required to run the cryptosystem at run-time, in addition to the plaintext or ciphertext.

109

## History

Abū Yūsuf Yaʿqūb ibn Isḥāq **al-Kindī** (801-873 CE) wrote the earliest known book on cryptanalysis, including details of frequency analysis (using aspects of probability some 800 years before Pascal and Fermat developed the topic). He also played a major part in introducing Indian numerals to the West.

http://en.wikipedia.org/wiki/File:Al-kindi.jpeg

http://en.wikipedia.org/wiki/File:Al-kindi_cryptographic.gif

110

## History

Mechanical tools have long been available to help avoid human error

http://en.wikipedia.org/wiki/File:16th_century_French_cypher_machine_in_the_shape_of_a_book_with_arms_of_Henri_II.jpg

111

## Modern uses


- Privacy
- Authentication
- Security
- Identification
- Message Integrity

http://commons.wikimedia.org/wiki/File:Smartcard2.png

112

## Encryption: Scytale

http://commons.wikimedia.org/wiki/File:Skytale.png

113

## Encryption: Caesar cipher

Replace each letter with the letter a fixed number of characters further on in the alphabet.

| | |
|---|---|
| Plaintext | ONCE UPON A TIME |
| Ciphertext | RQFH XSRQ D WLPH |

This system is called a **substitution cipher**; the **key** is the number of places in the alphabet that we need to move sideways.

We always substitute the same ciphertext character for the same plaintext character. This makes the system crackable using **frequency analysis**.

114

## Frequency analysis

WHEN DOES THE PLANE LEAVE?
**ZKHQ GRHV WKH SODQH OHDYH?**

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| plain text | 2 | | | 1 | 6 | | 2 | | | | 2 | | 2 | 1 | 1 | | | | 1 | 1 | | 1 | 1 | | | |
| cipher text | | | | 2 | | | 1 | 6 | | 2 | | | | 2 | | 2 | 1 | 1 | | | 1 | 1 | | 1 | 1 | |
| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z |

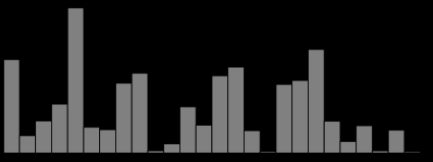
How often does each letter occur in the plaintext and ciphertext? The pattern of frequencies is the same in both rows, except that those for the ciphertext have been displaced 3 characters to the right. So this is a Caesar cipher with key 3.

115

## Frequency analysis (more generally)

http://commons.wikimedia.org/wiki/File:Dancing_men.png

What are the symbols? How often does each symbol occur in the message? Finding the frequencies helps us to guess which symbol



116

## Basic encryption: Vigenère cipher

- Named after Blaise de Vigenère, who described a version in 1586.
- Actually due to Giovan Battista Bellaso (1553).

http://en.wikipedia.org/wiki/File:Vigenere.jpg

This shows the encryption of HI THERE. The first E becomes W, but the second becomes E. This stops the system being cracked using frequency analysis.

117

## Decryption

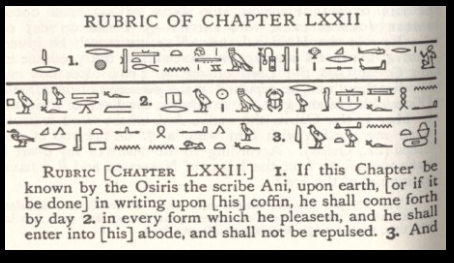
Can you obtain the plaintext corresponding to some ciphertext? If so, you can try working out how one was converted into the other, e.g. by identifying specific words or phrases.

Do you know what system was used to encrypt the ciphertext? If so, this will narrow down the number of options you need to consider.

Do you know what language the original message was written in? If so, you may be able to use properties of the language to help you decrypt messages (e.g. word endings, frequency analysis).

118

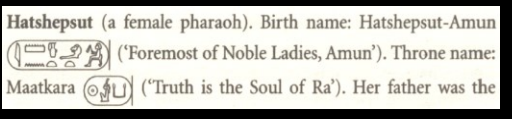
## How do we know what this says?




E. A. Wallis Budge. *The Book of the Dead*. New York: Gramercy Books (1960; originally published 1895)

119

## Identifying correspondences




Names of pharaohs are often written in oval boxes called cartouches. Comparing these against other languages tells us

120

## Further comparison with other languages shows…

- Hieroglyphs evolved from pictograms (pictures of things)
- Symbols can stand for one letter, two letters, three letters, or concepts
- The same symbol can have different interpretations at different points in the same document

http://commons.wikimedia.org/wiki/File:Papyrus_Ani_curs_hiero.j



## Using language properties

**Linear B was another script that no-one could read. The breakthrough was made by Michael Ventris in 1952.**

- **Ventris noticed that certain symbols kept re-appearing, and guessed that they represented male and female word-endings.**
- **He also noticed that some signs commonly occurred at the end of words in lists of (presumably) quantities of things, and deduced their meaning.**
- **In both cases, he used his knowledge of how grammar works in different languages. For example, in Latin you can say "and" by adding "*-que*" to the end of a word.**

http://en.wikipedia.org/wiki/Linear_B

**More information**: John Chadwick. *The Decipherment of Linear B*. CUP (1958)



## Uncracked ciphers still exist

- There are still languages that cannot be read
- There are even modern ciphertexts which cannot be read (except by their creator).

http://en.wikipedia.org/wiki/Phaistos_Disc



## Still a mystery

http://en.wikipedia.org/wiki/Voynich_manuscript



## The CIA Kryptos Sculpture

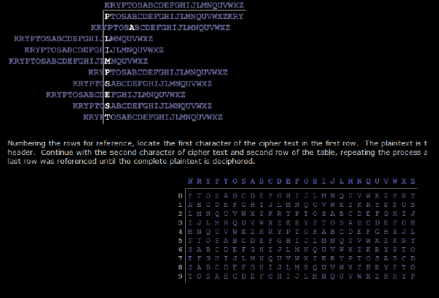
http://commons.wikimedia.org/wiki/File:Kryptos01_1.jpg

**Credits**: Picture provided to Wikimedia Commons by the sculptor, Jim Sanborn.



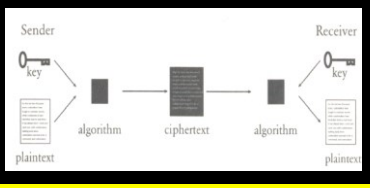


## Kryptos solution (panels 1-3)



http://realmoftwelve.kryptos.info/speculation/k1solution.html

## Private-key (symmetric) systems

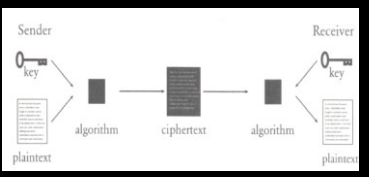



The sender and receiver both know and use the same key. But how do they share the key in the first place? If they have a secure channel for sending the key, why not

Picture from: Simon Singh. *The Science of Secrecy*. Fourth Estate (2000).

127

## Public-key (asymmetric) systems




The sender and receiver use different keys. Having one key doesn't provide enough information to deduce the other.

Picture from: Simon Singh. *The Science of Secrecy*. Fourth Estate (2000).

128

## Public-key cryptosystems

**Diffie**: This photo was provided to Wikimedia Commons by Mary Holzer from the Sun Microsystems Public Relations Department (April 2, 2004). **Hellman**: Wikimedia Commons (see http://en.wikipedia.org/wiki/File:Martin-Hellman.jpg for terms of reuse).

- Cocks (GCHQ 1973, disclosed 1997).
- Diffie & Hellman (1976)
- Rivest, Shamir, Adleman (1977)

129

## The RSA Cryptosystem

US Patent expired September 2000.

- Choose primes $p$, $q$
- Compute $n = pq$
- Compute $\phi = (p\text{-}1)(q\text{-}1)$
- Choose $e$ where
  - $1 < e < \phi$
  - $e$ and $\phi$ have no common factors
- Find $d$ such that $de = 1 \pmod{\phi}$

Publish the values $n$ and $e$. These form the PUBLIC key. The value $d$ is a PRIVATE key known only to the receiver.

- To encrypt the message m, compute $c = m^e \pmod{n}$. Send $c$.
- To decrypt c, compute $c^d \pmod{n}$.

This equals $m$ again!

It is HARD to compute $d$ if you are told only $n$ and $e$. There is no known (standard) algorithm that can do it quickly.

http://www.usc.edu/dept/molecular-science/RSA-2003.htm

130

## Secure communications via RSA

- A wants to send some message m
- A looks up B's encryption method E
- A already knows his own decryption method d
- A sends the ciphertext d(E(m))
- B receives d(E(m))
- B looks up a's encryption method e
- B computes e(d(E(m)) ==> E(m)
- B knows his own decryption method D
- B computes D(E(m)) ==> m

- The message is successfully transmitted.
- Only A could have sent it (since it is decrypted using e, and only A could have applied the corresponding algorithm ,d)
- Only B can read it (since was encrypted using E, and only B knows D)

131

## PGP (Pretty Good Privacy)

- Philip Zimmermann, 1992.
- Uses a combination of different techniques
- Considered to be extremely secure.
- Leaked across US border ("munitions" investigation)
- Source code published as a book (protected by First Amendment)

http://en.wikipedia.org/wiki/Philip_Zimmermann

132

## PRACTICAL CHALLENGE
## CRACKING THE VIGENÈRE CIPHER

133

### Groups of 2-4

1. Solve a puzzle to find the URL of your main challenge
2. Use the Vigenère cipher to encrypt a message
3. Submit it as a challenge and obtain a challenge of your own – another ciphertext
4. Try to work out the original ciphertext
5. You can use web tools to help you if you can work out what to look for

134

### Example: How would you decipher this?

The original is in English, a Vigenère cipher has been used, and the punctuation has conveniently been left visible.

```
ALPCS'Z RZ YSLH EZ ZPKSE O UMRSH SMRSH
VR L WWNLE YWNLE WWRI EZBPKSE,
TVV L YWNLE WWNLE'D PBX L DZPKSE ZPKSE
CU E WTUOX YTUOX WTYL XZYWNLE.
```

135

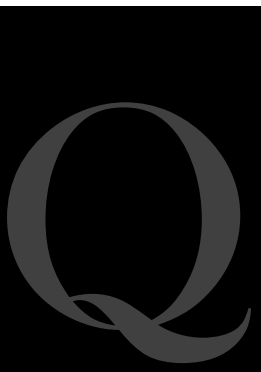
### Where to start

WKLV LV WKH SDVVZRUG BRX ZLOO QHHG:
SDVVZRUG

Open the file "handlingdata.pdf" at
http://staffwww.dcs.shef.ac.uk/people/M.Stannett/ambassadors

**Feedback sheets**: Please hand them in before you leave.
Don't forget to write your name on the sheet!

136



Any questions?

137