empath-e

MAKING CRM MAKE SENSE

Email Confidence Check
### Installation Instructions

---

CONTENTS

1.0 OPERATIONAL THEORY

As described previously, the ability of the utility is to ensure that all emails stored in your database are identified and checked for a "Confidence Level". No matter how thorough we attempt to be, there is no method that is 100% guaranteed to check an existing email for delivery – without actually sending an email and awaiting a response. As you can imagine, to do this would annoy the very people you are trying to check (and market to) – as they would physically receive an email from you.

Some email servers are setup to deliberately hide whom they deliver email to, and others simply accept email and pass it on to other servers for physical delivery. This is why we can only establish a **confidence level** on email addresses that appear to validate OK – but we can definitely identify the positively bad email addresses (badly formed email addresses, text/numbers added wrongly by users, non-compliance with standards etc) which probably form the bulk of your email addresses.

Therefore, given the above, the following shows how we establish and process email addresses:
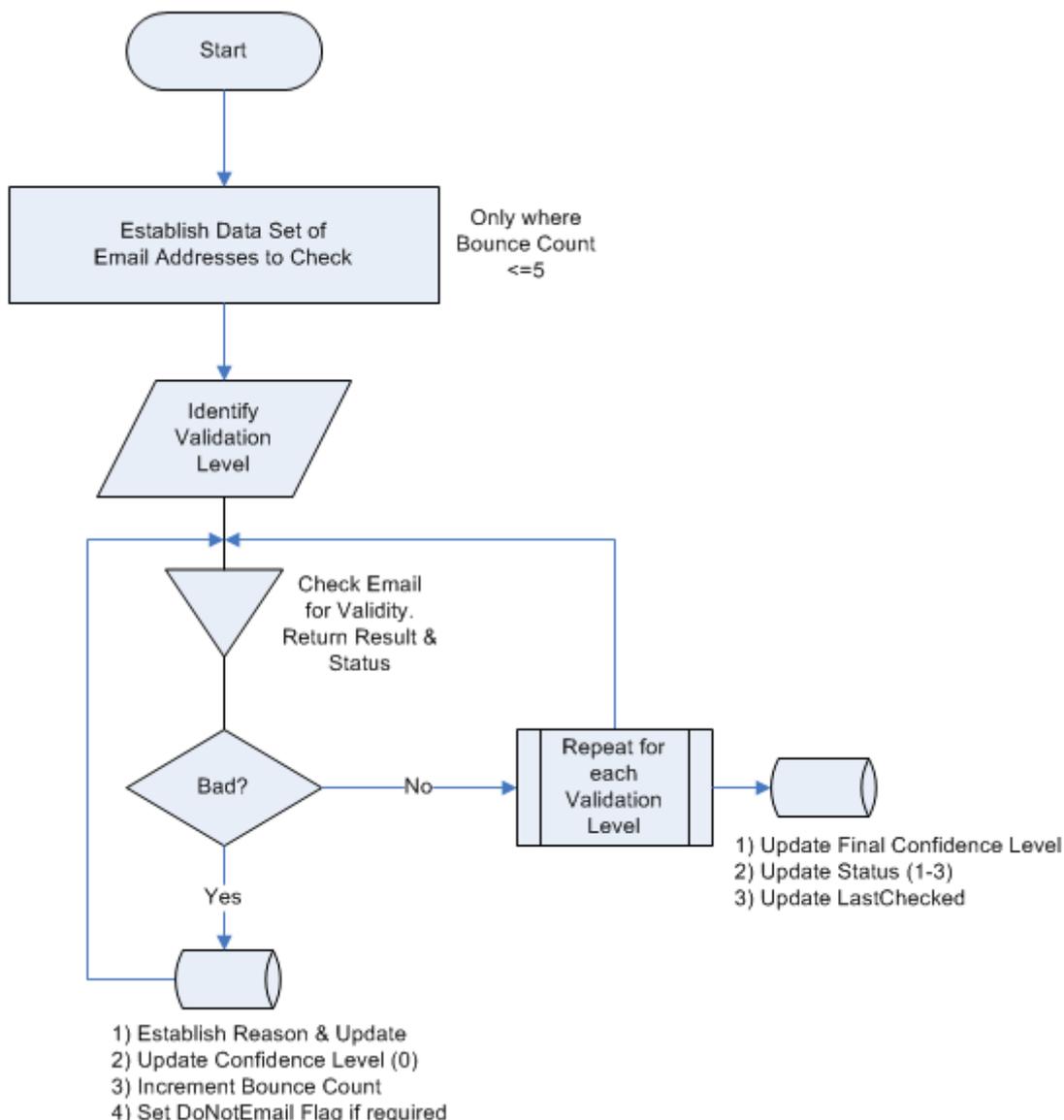
By following the process above we can establish the **Confidence Level** – but, as described before, this is <u>only</u> an indication.

Levels of Confidence

The first step of the utility is to establish just how far to check an email address. We have three levels for which to check each email address and each is more intensive that the last. For example, if Level 2 is requested then Level 1 is performed first and **only** if that is successful is Level 2 then performed.

Level 1 – Syntax

This will only check whether the email is syntactically correct and satisfies RFC 822 compliance. It is very quick to run (it does not need to contact any network servers) and can process around 90,000 emails in roughly 3 minutes.

Level 2 – DNS/MX

This level will perform a syntax check (level 1) and, only if that succeeds, will it then go on-line to check that a valid email server is known for this domain.

For example: mike.spragg@empath-e.com – this passes the syntax check. A lookup is then performed to establish what mail-server is defined as accepting email for empath-e.com (the domain). If either the DNS or MX lookup fails then the test fails. As it needs to identify the server (but doesn't connect to it) there is an inherent delay with this process. By default, the utility will time-out after 5 seconds.

Level 3 – SMTP

At this level – Level 1 and 2 are performed prior to this test. Only if that is successful does it proceed to perform a Level 3 test. This test involves physically connecting to the mail server responsible for receiving email for this domain. Commands are then sent to check that the mail server is responding correctly. Additionally, the utility attempts to find out whether the mail-server can receive email on behalf of the actual recipient. A "Success" is recorded should this be the case. Most email servers can respond positively that they can accept email for a particular mailbox – but, unfortunately, others simply respond "blindly". The result is still a "Success". Where a mail-server positively responds that it will not accept email for a recipient (left company, mailbox full etc) then this is recorded.

Due to the nature of this test, and depending on the number of emails to be checked, this is the most intensive test but will also take the longest to perform. You can set the SMTP Timeout to reduce the time taken to perform the test however, do not lower it so far that a slow email server cannot respond in time. This will lead to erroneous results.

It is suggested that you perform this test once and then revert to level 1 or 2 from then onwards (perhaps reverting to a full test every 3 months or so for example).

Bear in mind also that a mail-server can be off-line and mail-boxes do become full – so a positive failure today <u>may not be so in a week's time</u>. For each record checked, a counter is incremented and, when it reaches 5 (or a level you decide) then the email is recorded as permanently bad and removed from subsequent checks. 5 checks is considered an industry standard before assuming an email address is really out of action.

Interpreting Results

As we have seen – by setting the level we want to test and recording the results based on the result received we can establish the probability of an accurate email address. This can be summarised as follows:

	Confidence Rating (Highest successful validation level)			
	0 – Bad	**1** – Syntax	**2** - DNS	**3** - SMTP
Address is definitely bad	Y			
Syntax is OK		Y	Y	Y
Domain Exists			Y	Y
Domain has a working server				Y
Mail server did not reject email address				Y

Examples:

Check for Level 1 (Syntax) – the result can either be 0 (Bad) or 1 (Syntax OK)
Check for Level 2 (DNS) – the result can be 0 (Bad), 1 (Syntax OK) or 2 (Domain Exists)
Check for Level 3 (SMTP) – the result can be 0 (Bad), 1 (Syntax OK), 2 (Domain Exists), 3 (Valid Recipient).

See **Appendix A** for status text you can use to determine groups, queries in your database.

2.0 CONTENT

Requirements

Prior to running the utility ensure you have the following as a minimum:

- Windows XP, 2000, 2003 or later
- TCP/IP networking installed
- A directly-addressable DNS server (not through a proxy)
- Firewalls will need to allow outgoing traffic to UDP port 53 (for DNS) and TCP port 25 (for SMTP)

We suggest that you read this document entirely prior to configuring your system.

What's Supplied

We supply everything you need to get started. In the software distribution you will find the following:

em_settings_email.txt	The file required to configure your database connection and parameters
empath-e.EmailCheck.exe	The executable to run – either manually or via Windows Task Scheduler. Ensure you set the working directory to the directory you place the exe in.
empath-e Email Confidence Check.sxb	This is a bundle for Infor CRM – it is not required for other databases. It provides a form and the database schema changes required to operate the utility.
RegDLLs.cmd	A command file to run to register the relevant DLLs
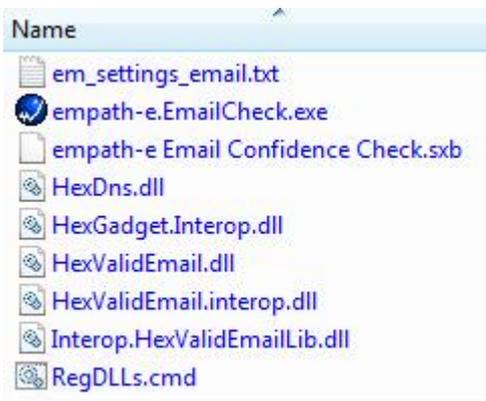
The Hex files are the components that have been sub-licensed from Hexillion. **Note:** You cannot use these for any other applications, as the run-time key is unique to the **empath-e.EmailCheck.exe** program.

Bundle Contents (Infor CRM Only)

The utility is pre-defined for Infor CRM 7.0.x or above. However, due to the generic nature of the utility you could use this for any relational database – it simply needs a valid connection and you have access to the SQL Code used in order to select and update the database.

The following plugins are supplied or updated in this bundle.

Type	Item	Description
COLUMN	CONTACT:EM_EC SMALLINT	Email Confidence Code (0-3)
COLUMN	CONTACT:EM_ESTATUS VARCHAR(64)	Email Return Status Text
COLUMN	CONTACT:EM_ELASTCHECK DATETIME	Last Checked Date (updated per run)
COLUMN	CONTACT:EM_ECNT SMALLINT	Bad Count (Incremented per run)
INDEX	CONTACT_EM_ESTATUS on CONTACT (EM_ESTATUS)	An index to assist performance
GROUP	EM_GoodEmails	Sample group of current good emails
GROUP	EM_BadEmails	Sample group of current bad emails
FORM	Contact:Email Confidence	Sample form to display above

If you are using a database (such as SageCRM or Microsoft Dynamics) you should create the same columns as indicated above. If you cannot do this – be sure to modify the SQL to cater for the column names you do create (see later in this document).

Impact Analysis (Infor CRM Only)

None of the additions will alter the system significantly in any way and can be easily removed without issue. By installing the bundle provided – all required columns are created for you automatically.

3.0 INSTALLATION

In order to start the utility you will need to copy in the files provided **and** modify the settings file. Ensure you do this **prior** to attempting to run the utility.

1) Create a new folder on the server/workstation you wish to run the utility on. This server/workstation will need access to the Internet (see requirements)

2) Drop the files supplied into this folder and double-click the REGDLLS.CMD file to initialise the components. **Note:** It assumes the folder contains the DLL's – you should modify this command file accordingly should you choose to move the files.

3) *Infor CRM Only:* Install the bundle empath-e Email Confidence Check.sxb via Administrator as indicated:

4) *Other Databases:* Be sure to create the relevant columns as indicated earlier/above. These are required in order to be able to update your system with the information coming back from the utility. Either use the same column names, or use your own – but ensure you then modify the SQL statements to match.

5) Open the em_settings_email.txt file and modify it according to the parameters described in the next section.

6) Once the settings are modified, you can now launch the application.

4.0 CONFIGURATION

The format of the em_settings_email.txt file takes the following parameters. Ensure you place the value immediately after the ":"

Connection	This is a standard OLEDB/ODBC connection string to your database. Contact support if you need details on what this should be set to. See Appendix B on how to define this string.
CheckLevel	1, 2 or 3 – this defines the level of check to run. Default = 1. 1 – Check syntax only 2 – Check Syntax then DNS/MX Records exist (slower) 3 – Check Syntax, then DNS/MX and then attempt a connection (slowest)
DNSTimeout	Default = 5000ms – can take a number between 0 and 10000. If you lower this figure too much, however, you may receive erroneous results for slow DNS queries.
SMTPTimeout	Default = 30000ms – can take a number between 0 and 60000. The largest number (60 seconds) will extend the time taken to run a complete check. Using these two numbers combined (35 seconds by default) is the maximum each email check could take.
FromDomain	You should enter your company domain e.g. empath-e.com as some servers may back-trace you to avoid spam techniques.
FromEmail	You should enter your email address (or postmaster@) should any diligent email administrator need to contact you. This will be logged in their system.
SetDNMail	Typically, Infor CRM only, this flag (Yes/No) will identify whether to update the standard column DONOTEMAIL to "T" – should an email fail to validate. ***Note:*** See UpdateDNEY, UpdateDNEN for examples on how to change this for your system
SelectStmt	This SQL statement identifies all the records to check. By default – it will only select those contacts with an email address where the "Bounce Count" is not over 5. As the process runs – the utility will increment this count automatically. Once it gets above 5 (generally considered as beyond soft-bounce) then the utility will never check this record again. You can override the count and behaviour as required. **Note:** only 3 columns are needed – the UniqueID, the Email address and the current count for delivery attempts. You should **not** modify this statement to have more than three columns defined and the order is fixed as shown.
UpdateDNEY	Depending on the flag SetDNMail (see above) then either this statement is executed, or the one below.
UpdateDNEN	By default, only the custom columns added are updated.

By default, the file is setup for an Infor CRM database, but you can modify it within reason to suit your own requirements.

As you can see – by using the replacement tokens {0}, {1} etc. you can choose your own statement to select/update your database.

5.0 RUNNING THE UTILITY

As the utility is designed to be run on the server, there is little to see. However, a form and progress bar is displayed. Additionally, the display updates with only the email addresses that fail validation (not all emails).

In the working directory a file called em_EmailStatus_Lvl_X.txt is also created with the output. X relates to the level of check being performed. You can check this for errors etc. once the utility has run. If you plan to run the EXE via Task Manager – be sure to set the Working Directory to the same folder as the utility and configuration file.

Viewing the results (Infor CRM Only)

If you login to Infor CRM you should find 2 contact groups called EM_BadEmails and EM_GoodEmails – these will contain the results of the check:

	DoNotEmail	Email Confidence	Email Status		Bad Count	LastChecked
	No	0	Unbalanced square bracket		3	26/03/2009 00:43
	No	0	Unbalanced square bracket		3	26/03/2009 00:43
	No	0	Unbalanced square bracket		3	26/03/2009 00:43
	No	0	No local part specified		3	26/03/2009 00:43
	No	0	No local part specified		3	26/03/2009 00:43
	No	0	No local part specified		3	26/03/2009 00:43
	No	0	No domain specified		3	26/03/2009 00:43
	No	0	No domain specified		3	26/03/2009 00:43
	No	0	No domain specified		3	26/03/2009 00:43
	No	0	No domain specified		3	26/03/2009 00:43
	No	0	No domain specified		3	26/03/2009 00:43
	No	0	No domain specified		3	26/03/2009 00:43

Here, we can see that several errors were found and the reason shown. The confidence is zero (bad email address) and the reason varies. In this instance, down to bad data in the email address. The count is also at 3 – meaning they have 2 more checks before they are no longer checked again.

By changing the level of checking to Level 3 – we can see different results are returned:

	DoNotEmail	Email Confidence	Email Status		Bad Count	LastChecked
k	No	2	Connection refused		1	26/03/2009 00:56
	No	2	Connection reset		1	26/03/2009 00:56
	No	2	Could not verify recipient		1	26/03/2009 00:56
	No	2	Could not verify recipient		1	26/03/2009 00:55
	No	2	Could not verify recipient		1	26/03/2009 00:56
A.CO.	No	2	Could not verify recipient		1	26/03/2009 00:56
	No	2	Could not verify recipient		1	26/03/2009 00:56

We are seeing Level 2 results (we requested level 3) – and this is down to various errors with the mail-servers involved. Notice that where an email cannot be verified it is **not** set to Bad (0). This is because there could be a number of reasons why this cannot be reached right now.

A new tab is also available – this shows you the email address and its status. Your business partner can use this form, perhaps as a template to copy the relevant details onto your main contact view if required.

If a user discovers that an email address has been incorrectly rated – they can use the button "**Reset for Checking**" to eliminate the data recorded. This will put the contact into the queue for the next time the utility is run.

APPENDIX A – RESULT CODES

The following are the possible values returned when an email is checked for validity:

Timed out
Connection refused
Connection reset
Host unreachable
Address not available
Network down
Network unreachable
Connection aborted
Host not found
Try again
No recovery
No data
Unexpected error
Address too long
Extra text is present
Illegal character
Unbalanced parenthesis
Unbalanced square bracket
Unbalanced angle bracket
Quotation marks not closed
A domain literal is present
Misplaced domain literal
Misplaced quoted string
No local part specified
No domain specified
Invalid domain
Invalid domain literal
No DNS servers configured
Domain does not exist
No mail exchange for domain
Could not verify recipient
Recipient rejected
No address specified
Mailbox is full
MX records do not have valid IP addresses
SMTP server is blocking email from this sender
SMTP server does not reject bad addresses

APPENDIX B – DEFINING A CONNECTION STRING

In order for the utility to be able to connect to your database it needs to know how to do this. The connection string is simply a Microsoft Universal Data Link construct.

The simplest way to define the definition is to create a text file on your desktop called link.udl and then double click it. Complete the necessary configuration options (ensure you choose the "save password" option) and then rename the file to link.txt.

You can then open this file and copy the information into the Connection: parameter.
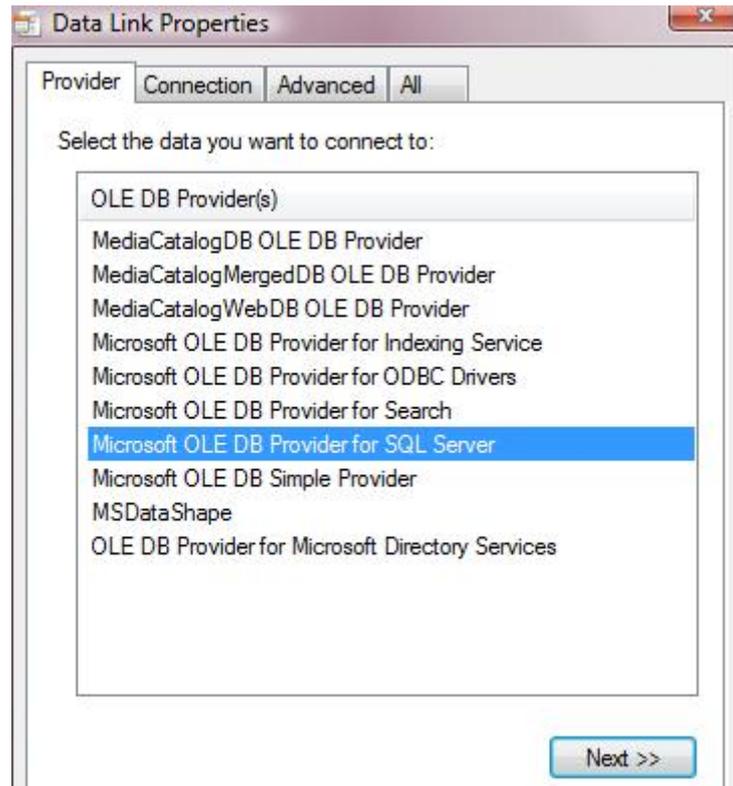
Note: You can choose any provider here, including the Infor CRM provider. You **must** use the Infor CRM provider if you want the data to synchronise and you are using remotes. If you use the Microsoft SQL Native Client – this can also be used in a remote situation, except the email status will not be synchronised to other offices/users (which is OK if you only email from one location).

The text file will look similar to the following:

Provider=SQLNCLI.1;Password=**{yourpassword}**;Persist Security Info=True;User ID=**{yourname}**;Initial Catalog=**{yourdatabase}**;Data Source=**{yourserver}**

{yourserver} – use (local) if running this utility on the SQL Server, otherwise use the server name.
{yourdatabase} – use the physical name of the database
{yourname} and **{yourpassword}** – are specific to your database type. Infor CRM always uses "sysdba" as the username and "masterkey" as the ordinary password.