Textual Integration -- Some Business Examples

ABSTRACT

For years organizations have been making decisions based on transaction processing data. There is nothing wrong with this. But there is another form of data that needs to be included in the corporate decision making process, and that type of data is unstructured textual data.

This presentation addresses some of the issues of decision making with textual data and ends with an in depth, real example, where the attendee can see how text can be reduced to a data base that can be queried and analyzed.

BIOGRAPHY

William H. Inmon
Inmon Consulting Services

Bill Inmon, is recognized as the "father of the data warehouse" and co-creator of the "Corporate Information Factory." He has 35 years of experience in database technology management and data warehouse design. He is known globally for his seminars on developing data warehouses and has been a keynote speaker for every major computing association and many industry conferences, seminars, and tradeshows.

As an author, Bill has written about a variety of topics on the building, usage, and maintenance of the data warehouse and the Corporate Information Factory. He has written more than 650 articles, many of them have been published in major computer journals such as Datamation, ComputerWorld, and Byte Magazine. Bill is currently a columnist with Data Management Review, and has been since its inception. He has published 45 books; one sold over half a million copies, 21 have been book club selections with publishers such as Prentice-Hall, John Wiley, and QED.

Translations of various books have been done in Chinese, Dutch, French, German, Japanese, Korean, Portuguese, Russian, and Spanish.

TEXTUAL INTEGRATION – SOME BUSINESS EXAMPLES

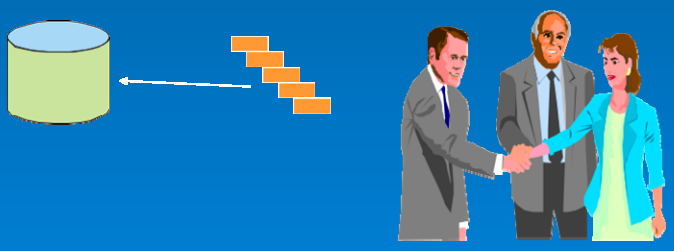
A presentation by
W H Inmon

© Copyright Inmon Consulting Services, 2008

in no small part, corporate decision making
has depended on data and data bases for years

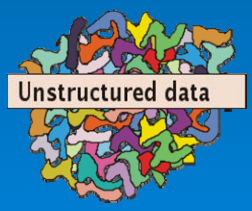
we take that for granted

© Copyright Inmon Consulting Services, 2008



and for years the data that has populated the data bases of the world has been based on business transactions –
- banking activity
- retail sales
- insurance claims
- telephone calls
- airlines reservations, etc

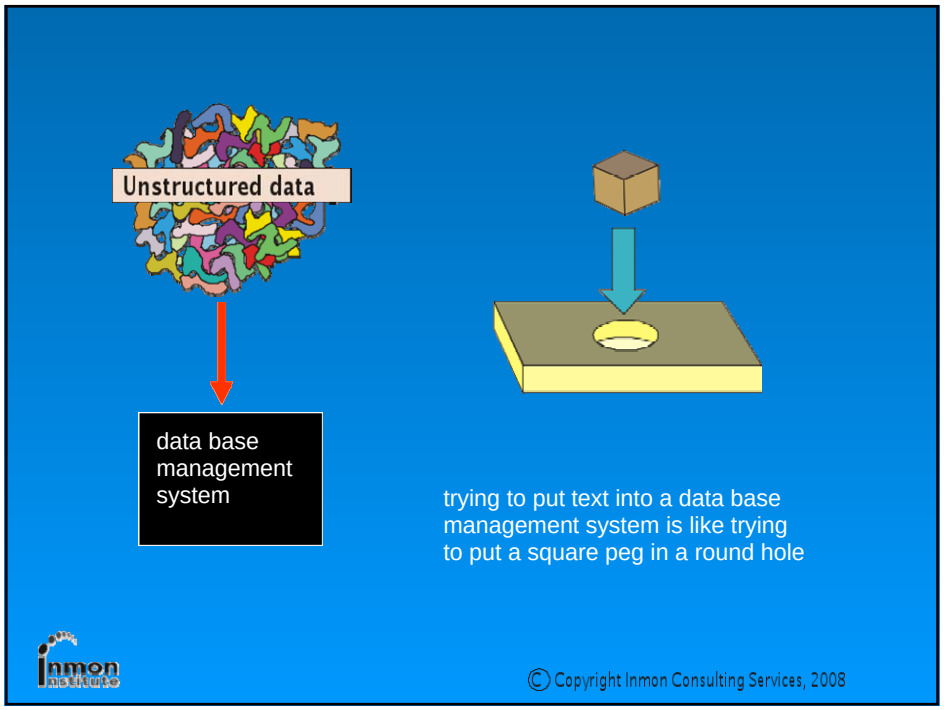
© Copyright Inmon Consulting Services, 2008

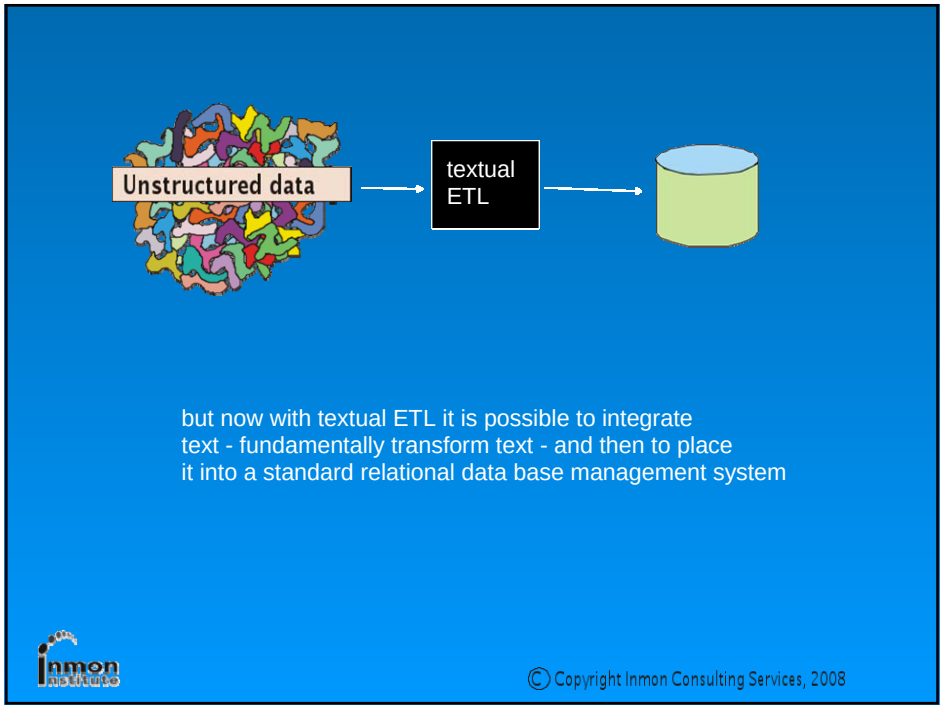


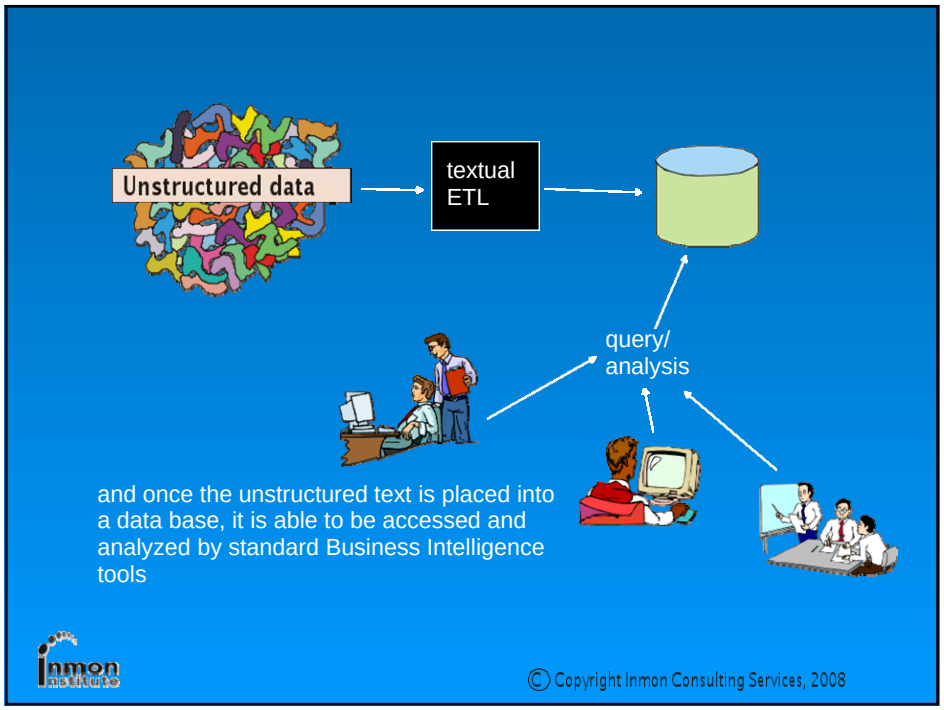
but there is another important type of data that has not been included in corporate decision making – textual unstructured data

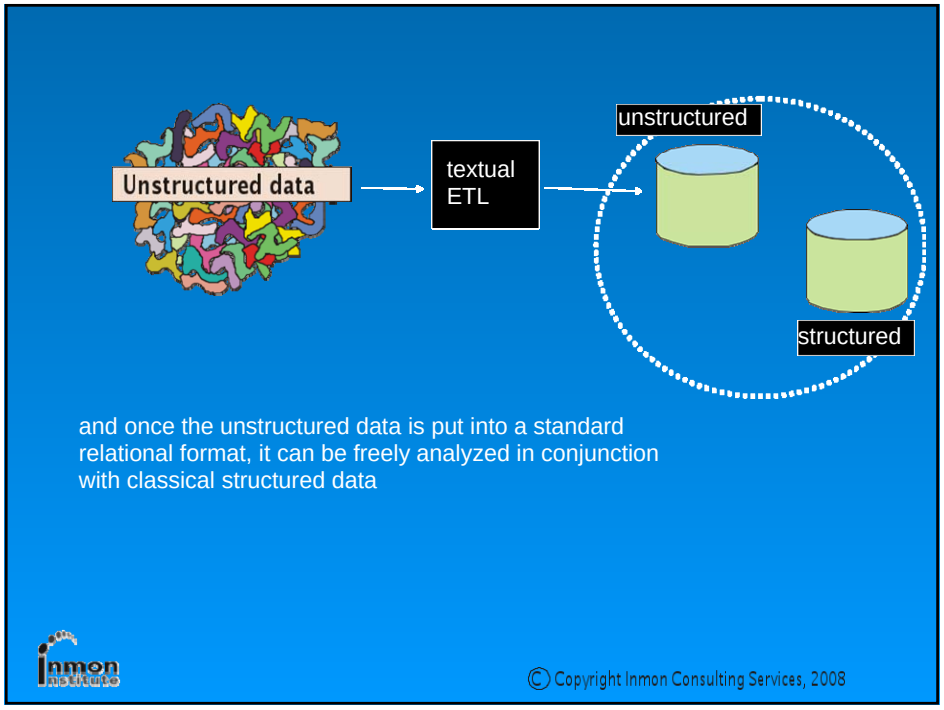
email
contracts
memos
warranties
letters
law suits
management reports
etc.

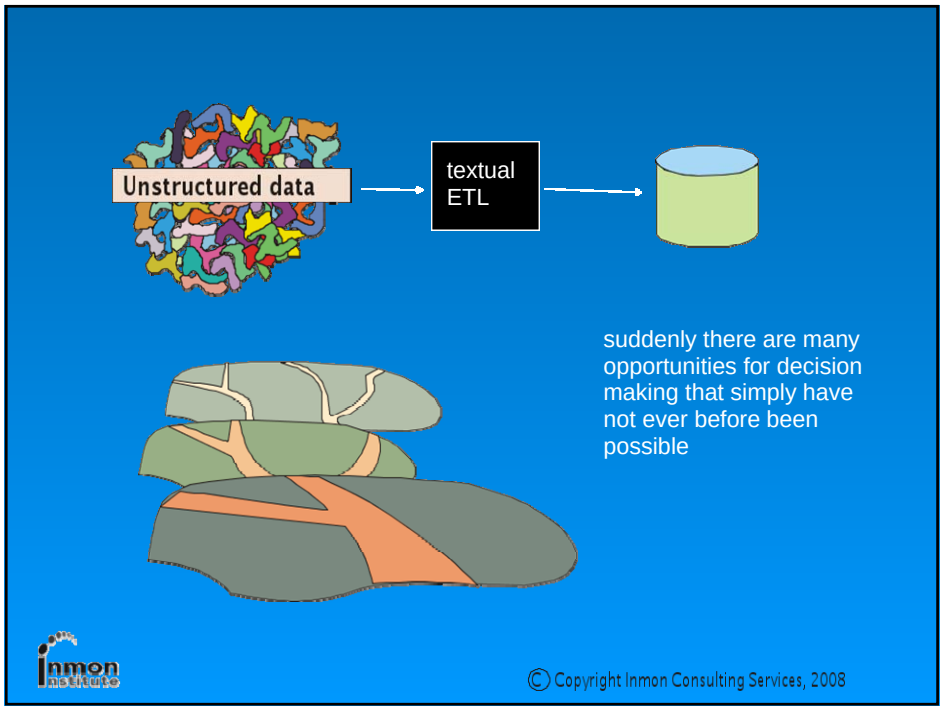
© Copyright Inmon Consulting Services, 2008

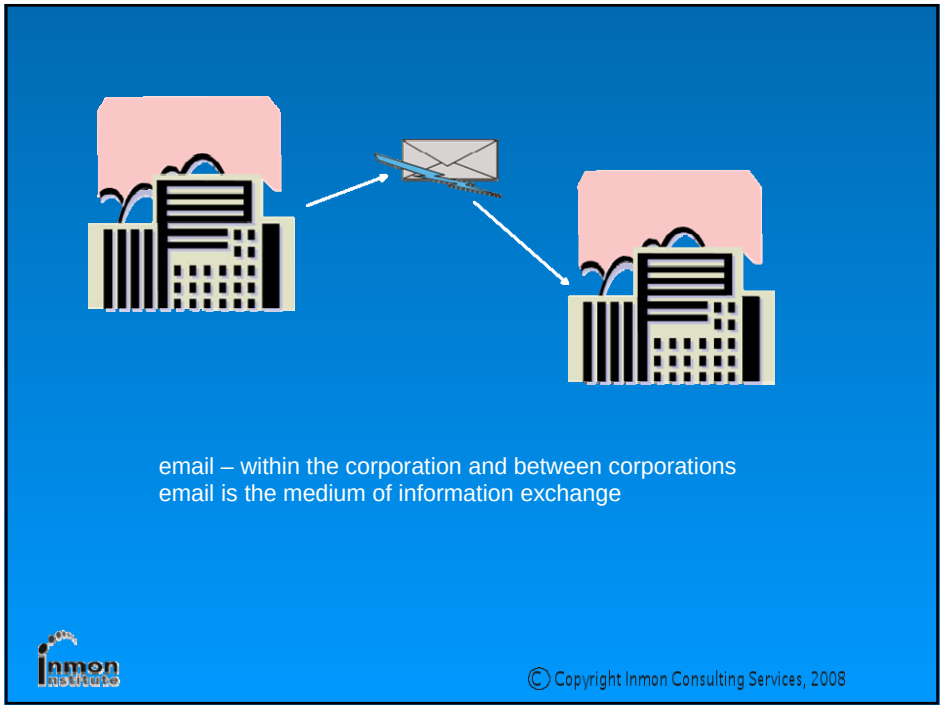

Unstructured data
data base management system
trying to put text into a data base management system is like trying to put a square peg in a round hole
inmon institute
© Copyright Inmon Consulting Services, 2008

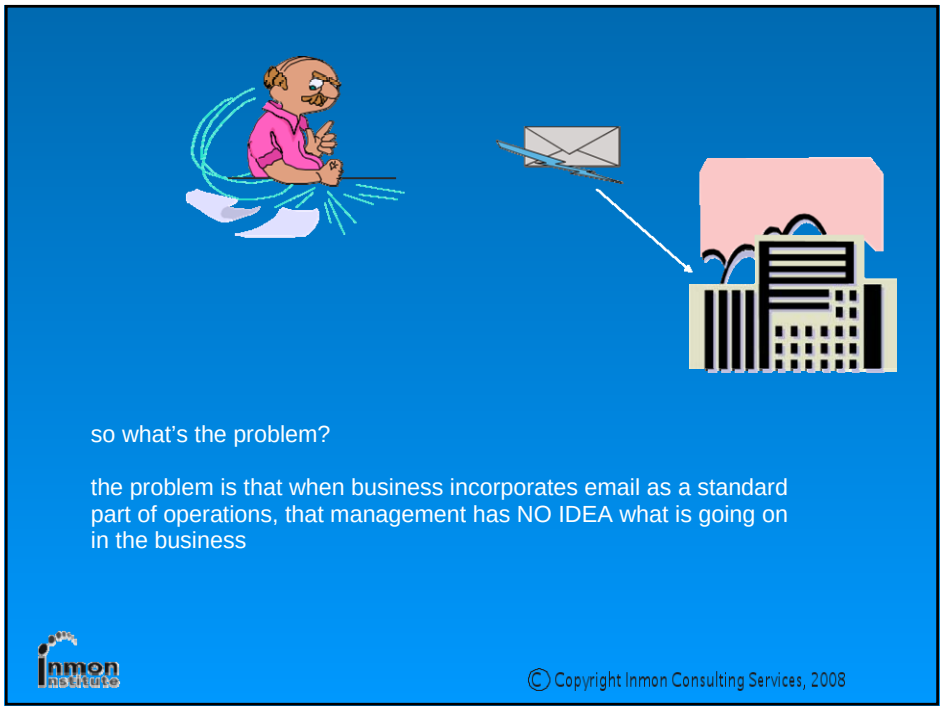

Unstructured data
textual ETL
but now with textual ETL it is possible to integrate text - fundamentally transform text - and then to place it into a standard relational data base management system
inmon institute
© Copyright Inmon Consulting Services, 2008


Unstructured data
textual
ETL
query/
analysis
and once the unstructured text is placed into a data base, it is able to be accessed and analyzed by standard Business Intelligence tools
inmon institute
© Copyright Inmon Consulting Services, 2008


Unstructured data
textual
ETL
unstructured
structured
and once the unstructured data is put into a standard relational format, it can be freely analyzed in conjunction with classical structured data
inmon institute
© Copyright Inmon Consulting Services, 2008


Unstructured data
textual
ETL
suddenly there are many
opportunities for decision
making that simply have
not ever before been
possible
inmon institute
© Copyright Inmon Consulting Services, 2008

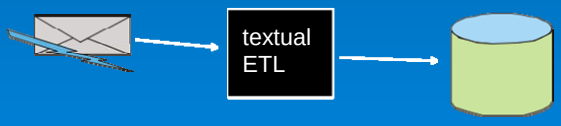

email – within the corporation and between corporations
email is the medium of information exchange
inmon institute
© Copyright Inmon Consulting Services, 2008



so what's the problem?

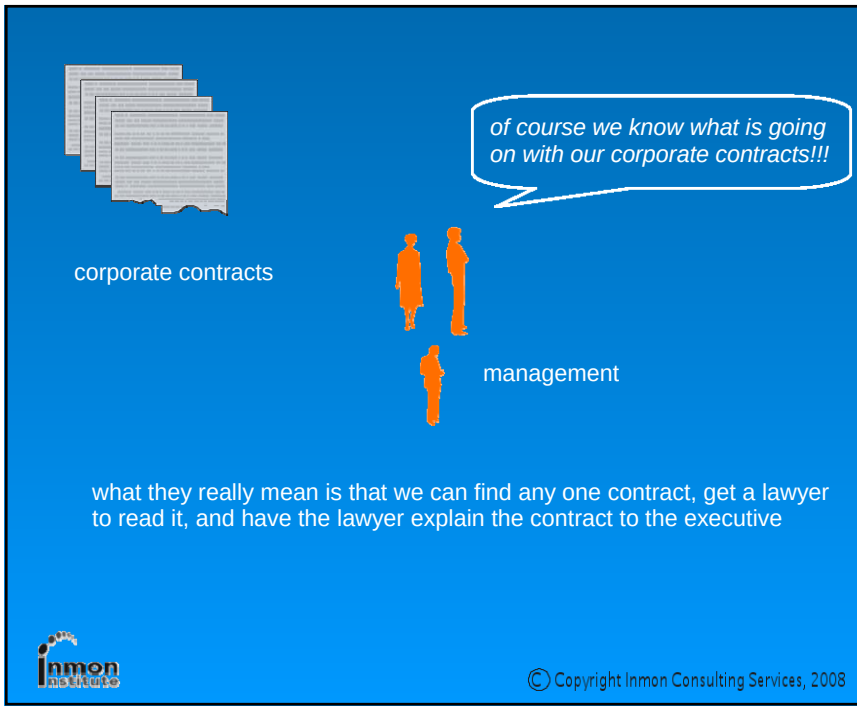
the problem is that when business incorporates email as a standard part of operations, that management has NO IDEA what is going on in the business

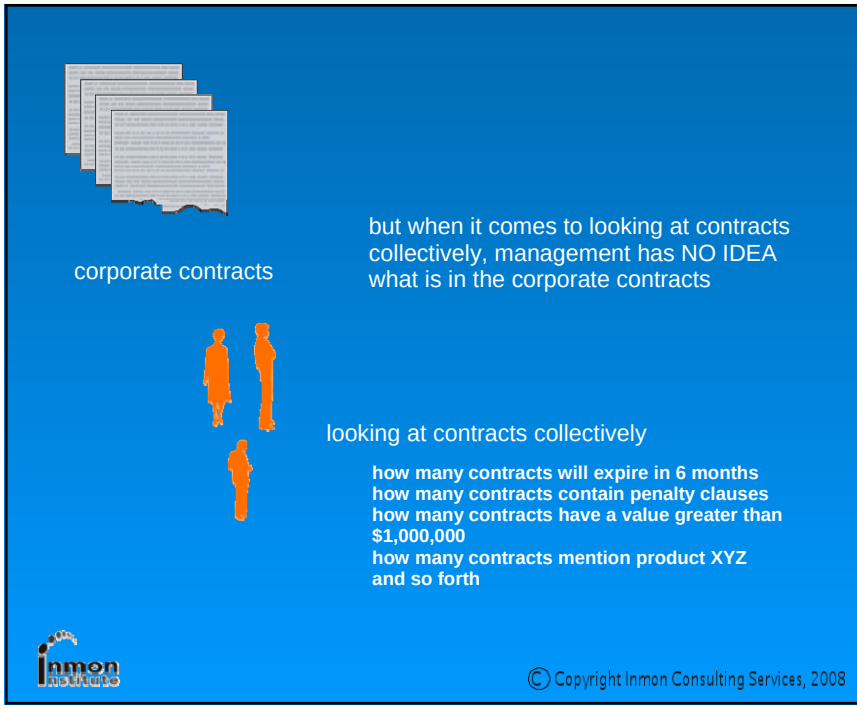
© Copyright Inmon Consulting Services, 2008

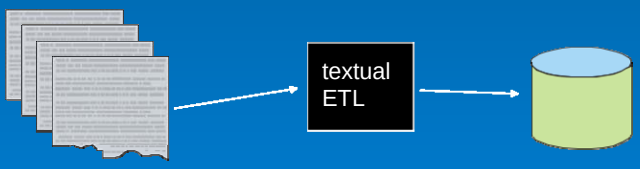


now management can see what is going on with the business

© Copyright Inmon Consulting Services, 2008


of course we know what is going on with our corporate contracts!!!
corporate contracts
management
what they really mean is that we can find any one contract, get a lawyer to read it, and have the lawyer explain the contract to the executive
© Copyright Inmon Consulting Services, 2008

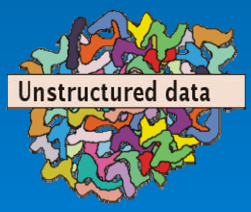

corporate contracts
but when it comes to looking at contracts collectively, management has NO IDEA what is in the corporate contracts
looking at contracts collectively
how many contracts will expire in 6 months
how many contracts contain penalty clauses
how many contracts have a value greater than $1,000,000
how many contracts mention product XYZ
and so forth
© Copyright Inmon Consulting Services, 2008



corporate contracts

now with corporate contracts you can look at them collectively

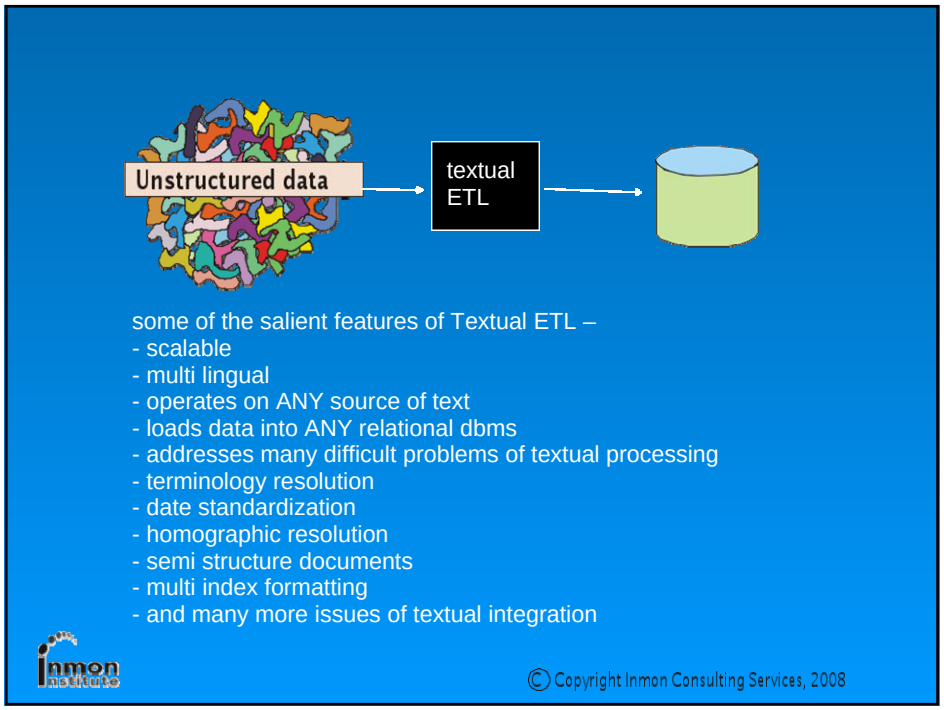
© Copyright Inmon Consulting Services, 2008



there are many other forms of textual information that are important to integrate into a standard relational data base –

- loans applications - insurance claims
- warranties - doctors notes
- chat sessions - telephone messages
- customer relationships - hr evaluations
- health care results - and many more

© Copyright Inmon Consulting Services, 2008


Unstructured data
textual
ETL
some of the salient features of Textual ETL –
- scalable
- multi lingual
- operates on ANY source of text
- loads data into ANY relational dbms
- addresses many difficult problems of textual processing
- terminology resolution
- date standardization
- homographic resolution
- semi structure documents
- multi index formatting
- and many more issues of textual integration
inmon institute
© Copyright Inmon Consulting Services, 2008

so now let's look at an example -
inmon institute
© Copyright Inmon Consulting Services, 2008