kompany

The Trusted Source for Official Company Information

Peter Bainbridge-Clayton
Founder and CTO

peter.bc@kompany.com
May 2015

Easy Access to Business Background Information

kompany is a 3 year old Vienna based company with the aim of providing ccess to global, official corporate information enhanced by other official d (e.g. Announcements, Patents, Trademarks), 3rd party researched data (e.g Credit Reports and Scores), and other data from a variety of sources, but always with a base of official data.

***W*e take information from these sources, consolidate, merge and enhance dding value in the process and then provide access through our multilingu multicurrency website (www.kompany.com) and our API (brex.io).**

his makes Open Data initiatives very interesting to us from a data access a commercial point of view. Often the data we want is inaccessible, accessi only through non-automated methods, or only available at exorbitant cost

fficial, legally correct information is the bedrock of our business model – i the basis we use for providing accurate, useful, timely and cost-effective information and services. Open data does, however, bring with it a set of

Easy Access to Business Background Information

kompany is NOT:

Associated with Vincent Kompany!
A yellow pages site
Just a reseller of credit reports

kompany IS:

An enabler for consumers, businesses, government agencies and departments to Verify, Evaluate and Validate
Themselves
Suppliers
Customers
Partners
Applicants

kompany aims to be:

The global clearing house for official corporate register data

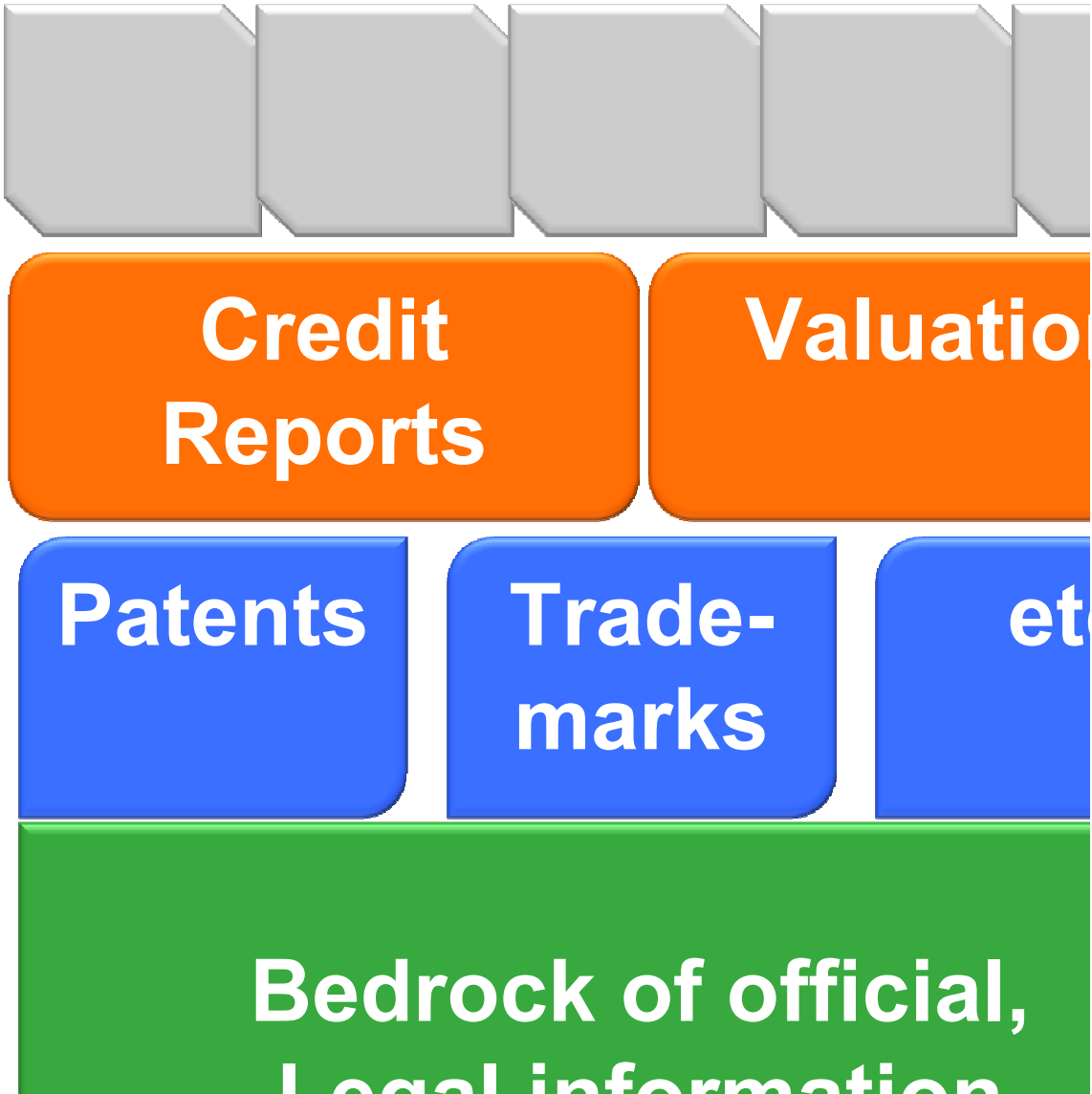
Easy Access to Business Background Information

Other sources (web, 3rd party, social media etc)

3rd Party researched Information

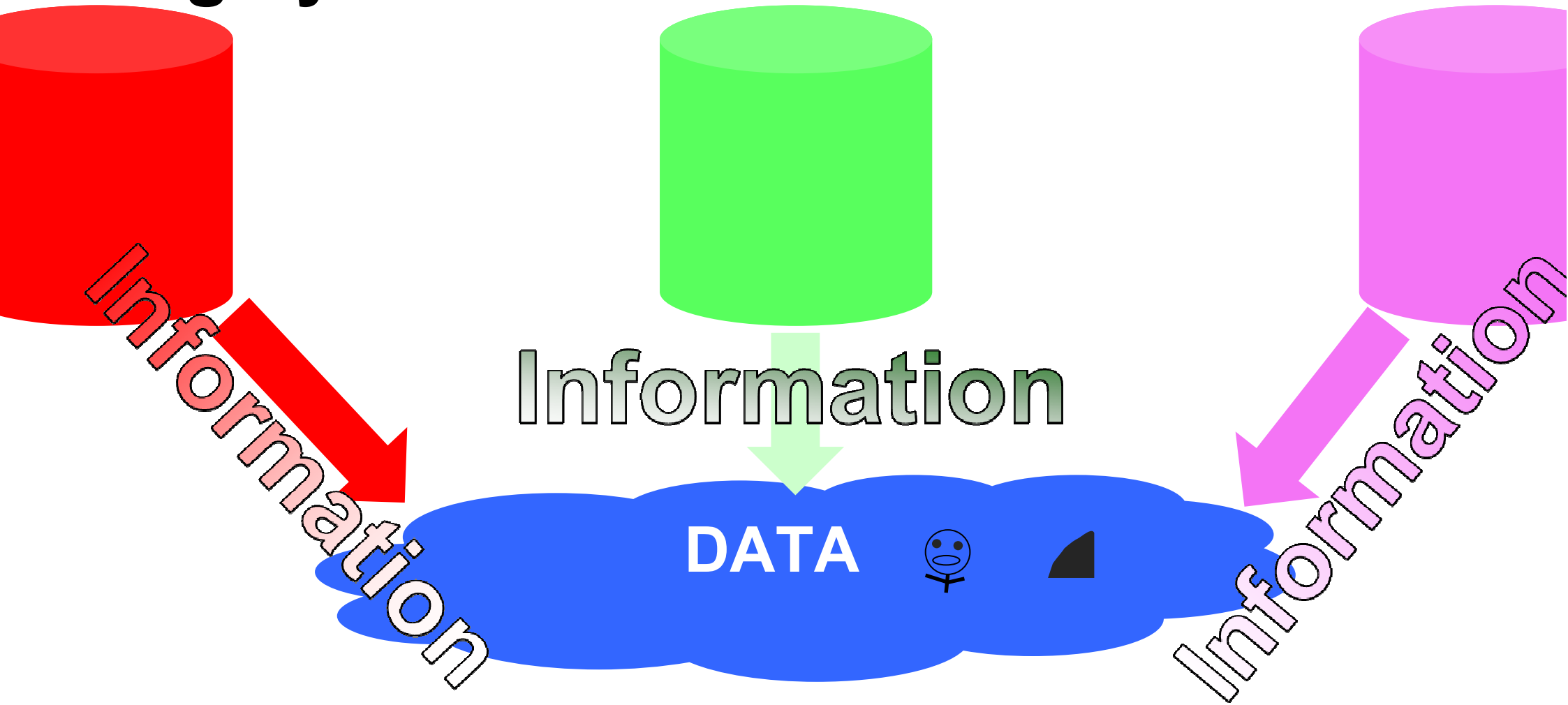
n-corporate register but official information

ect or indirect connections to orate registers of 28 countries, JS States with agreements in



Turning data back into information

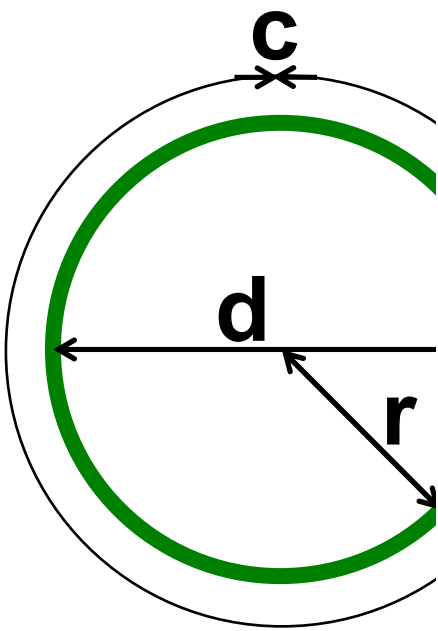
Highly contextual silos of information



king information out of a silo, we run the risk of losing the context, and a m
ation without context is just a sea of data – in the same way as a library wi

but Context is Key

$$c=2\pi r \quad \text{or} \quad c=\pi d$$



Lets add some context: GEOMETRY

"The ratio of a circle's circumference to its diameter is a constant called pi"

ss context: ~~GEOMETRY~~→MATHEMATICS

$$c:d = \pi \approx 3.141592654$$

en Less context: ~~GEOMETRY→MATHEMATICS~~→DATA

# of a company?

For an Austrian GmbH this will be a Geschäftsführer, which would translate in English to “Managing Director”

For a UK Limited this will be an Officer of type Director – which is not necessarily the same as a Managing Director, a Managing Director might run the company but not necessarily have the legal responsibilities of a Director. Managing Director is really just a job title in this context.

So – can we use Geschäftsführer on the German language version of our website when talking about a UK Limited company? Yes – because we know the expected context is the legally responsible person. As long as we show a UK Director as a Geschäftsführer we have successfully mapped the contexts and displayed the expected result.

Another pertinent example might be to have a credit score for a company of 5, without any idea of the allowed range. Is it 5 out of 5, or 5 out of 100?

What is context in this context?

he OED defines context as:

**The circumstances that form the setting for an event, statement or idea, ar
in terms of which it can be fully understood”**

/hich for our use could be given a more technical definition of:

**The data elements and relationships including their meaning and limits, th
pertain to an entity and in terms of which it can be fully understood”**

is important to recognise that context is not just a function of what information is
available, but is also altered by our personal expectations, knowledge and
preconceptions.

rieve → Analyze → Transform → Store → Enhance → Play

consistency between information providers on access methods, acc
tocols, data formats, authentication mechanisms, etc. Essentially ev
vider has a different interface, and for countries like the US, each St
potentially has a different interface

we built an extensible toolbox of interfaces including:

L over REST

AP

TPS scraping

JSON over REST

HTTP scraping

XML file

rieve → **Analyze** → Transform → Store → Enhance → Play

lthough there is some commonality in terminology and data structure
hese tend to be aligned along language and cultural lines in the case
countries and technical lines in the case of e.g. the US.

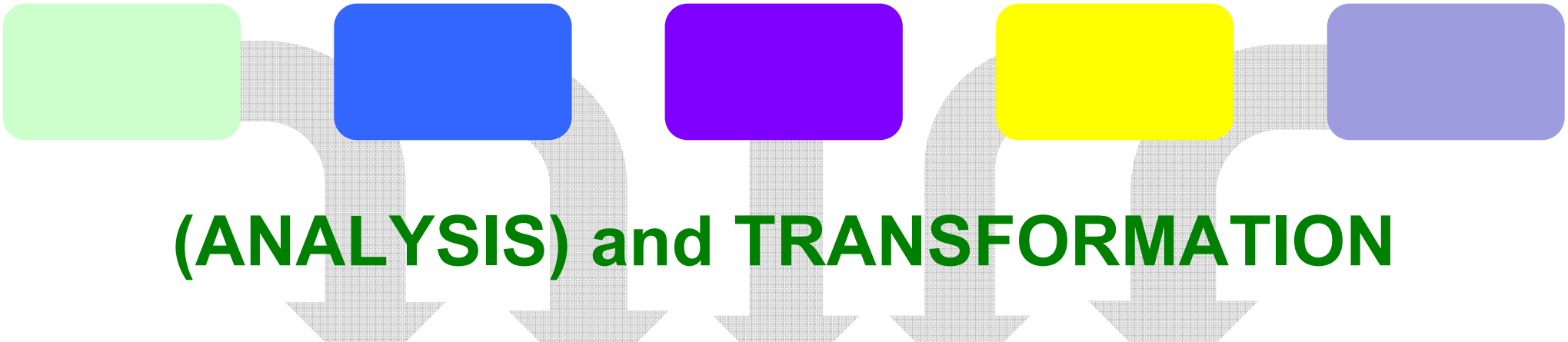
example, Austria and Germany have different registration structures
access methods, but the company types, their legal meanings and th
rminology is highly compatible, whilst the USA is in general based o
all number of technical implementations, each underpinning a numb
es, and each presenting very similar information in different ways thr
different methods.

This is very much a human exercise requiring good knowledge of th
guage, the legal frameworks and the terminology – the outcome of w
is a consistent set of terms which can be used internally.

rieve → Analyze → **Transform** → Store → Enhance → Play

e able to handle different data sources with different datasets, elements of which
ve the same or different meanings, it was technically imperative to create an inte
sentation which maintains meaning, maintains context, and allows for data corre
d enhancement. This mechanism also had to deal with unknown types of data a
unknown relationships between their elements.

chieve this we defined an extensible ontology onto which all data elements and
nnections are mapped – this is fed from the analysis step and allows us to maint
ontext across differing sources. Each of the interfaces has the task of mapping t
information it receives onto this ontology



(ANALYSIS) and TRANSFORMATION

rieve → Analyze → Transform → Store → Enhance → Play

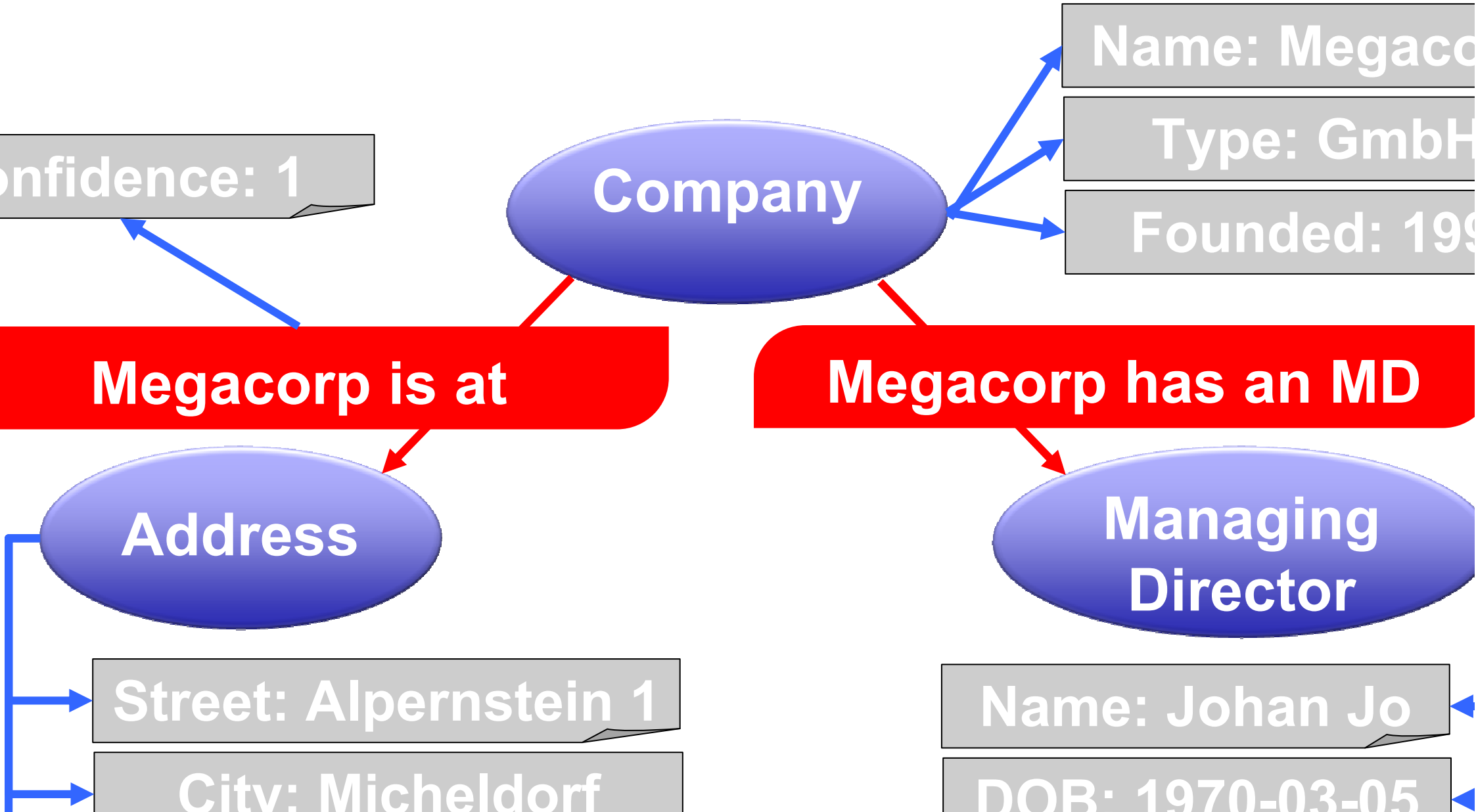
e storage problem is essentially the same as the transformation problem. We ha
ise a method of storing unknown types of data from unknown sources with unkn
eanings and unknown relationships in an open ended way which would not requ
changes to the data structure with every new source.

rtunately there was already a model which met these requirements published by
W3C – Resource Description Framework, RDF.

RDF basically provides an open, flexible model for describing entities, data and
tionships in a web friendly format (for identification and/or access), but also an R
compatible structure is used internally for our ontology and data storage.

coupled with domain knowledge and careful upfront analysis, allows us to store
perties and its relationships that are unknown until analysis time whilst still retai
context, and without rebuilding the data store. It also importantly provides forwar
npatibility with other RDF compatible applications, particularly those on the sema
web.

rieve → Analyze → Transform → **Store** → Enhance → Play



rieve → Analyze → Transform → Store → **Enhance** → Play

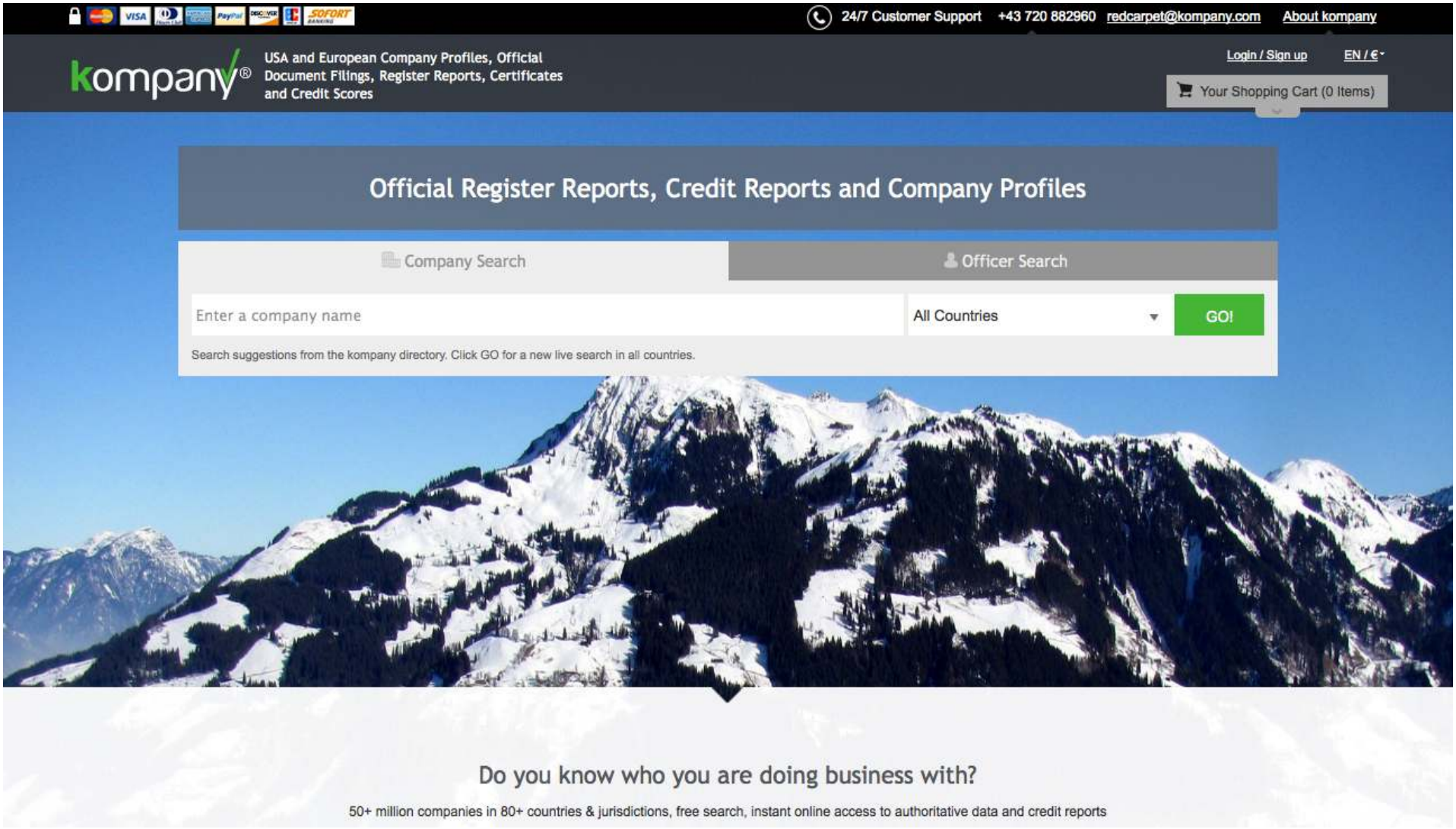
Structuring the data in this way allows us to use several techniques t
enhance the data.

‘s for example allow us to visit the website and possibly extract data
e impressum, or follow links to LinkedIn, Facebook, etc and extract d
from there.

e can also parse non-structured information in a programmatic fashio
extract meaningful and useful information, e.g. German Corporate
Announcements, Swiss Corporate Announcements

*W*e are also testing AI techniques (evolutionary algorithms and neura
works) to enhance and expand the data we have – by looking for patt
e information we already have to extract new relationships – for exa
etect when a person across multiple companies, possibly across mu

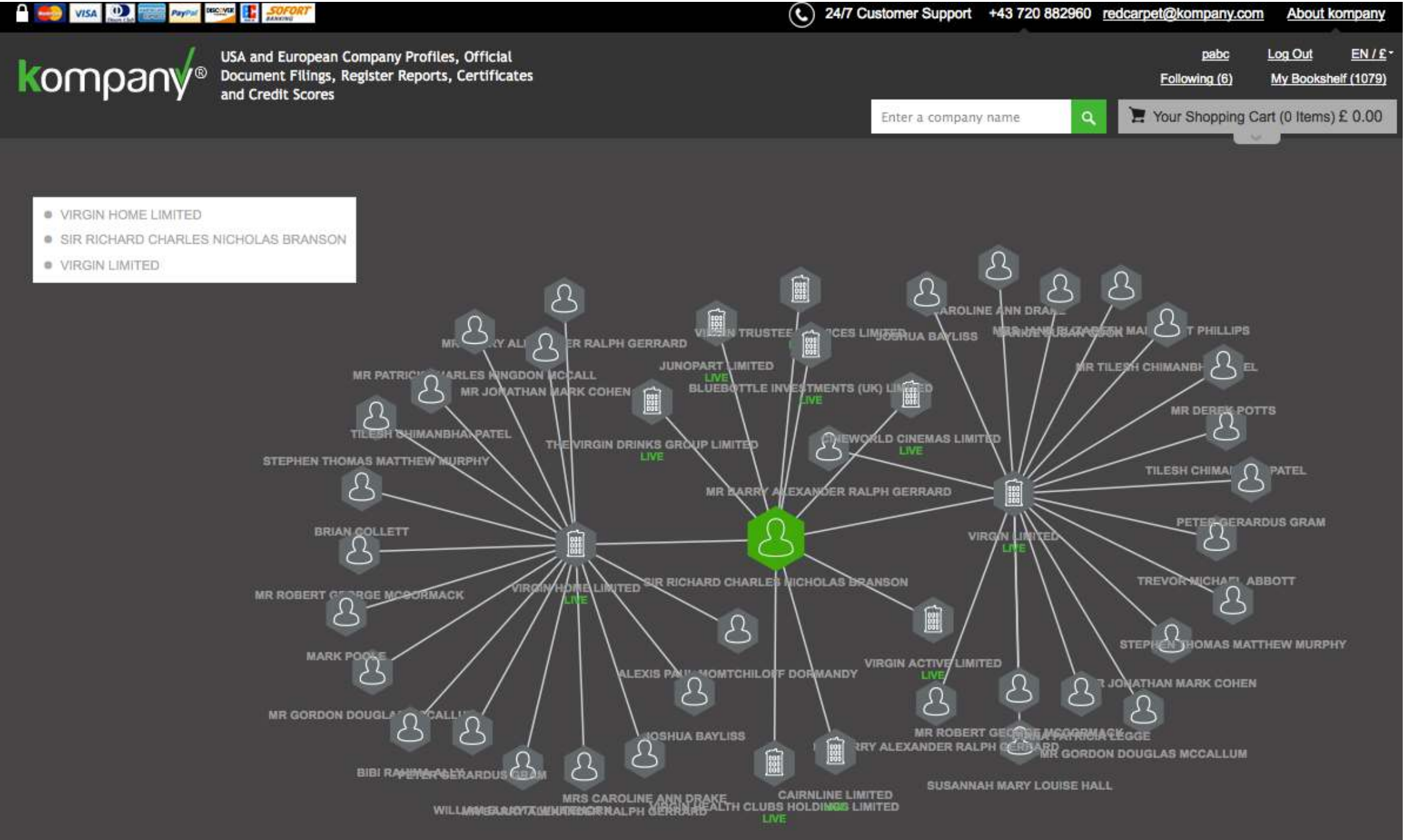
rieve → Analyze → Transform → Store → Enhance → Play



www.kompany.com

Multi lingual, multi currency, multi country company search, Information and products

rieve → Analyze → Transform → Store → Enhance → Play



You can build a company on public data, can you build a business?

At a simple level, a successful business needs:

Something to sell

Verify, Evaluate and Validate capability – KYC, KYS, KYB and AML

Someone to sell to

100 million firms trade locally, nationally and internationally across the globe

Good Suppliers

Low as possible costs

What is a Good Supplier of Public Company Register Data?

"A good supplier provides reasonable cost, automated, access timely, accurate, officially licensed structured data"

	Preferred	Get ☹ - ☺
sonable cost	Realistic charges < 50c /doc	5million Euro –
nated Access	Webservice, XML/JSON	XLS files – Webscrap Webservices
ly, Accurate	Hourly update with notification of change	Posted DVD's – Mo Files – Hourly Unstructured – Daily
		XLS files, CSV files,

Officially Licensed

ere are significant questions still to be addressed in terms of PSI for corporate d
specifically around:

wnership – does the data belong to the company, the register, the public or no-c

pyright – does copyright exist with the register, the company or does it exist at

-Use – if we take data and enhance it, or use it as the basis for enhanced produ
s that reuse, or is it a derivative?

igin of Source – is there a requirement to maintain an audit trail of initial data
source? What if the data is superseded?

vacy – Is there a right to privacy around data which is already publically access
officially' published, or as a derivative?

early these are interlinked, and until these are addressed at a supplier level (i.e a
egislative level), the onus remains on the user of the information to decide on th

but there is a high risk of making lawyers rich!

Thank You

here are all these companies making millions from our data so they should pay for that

All our data will be available free of charge by the end of the year