The irresistible rise of White Box Data
Chris Taggart
CEO & Co-Founder of OpenCorporates, the largest open database of companies in the world
Once upon a time....
Data was...
- local, siloed
- expensive to store
- just good enough
...and **external** data was Black Box

- no clear definition
- no provenance
- proprietary identifiers
- limited utility
- built for the analogue, single record, one-hop world
The Black Box model (simplified)

Original base data
- Provenance
- Local identifiers
- Clear model

Black Box data
- No Provenance
- Proprietary identifiers
- Opaque model
The Black Box model (in reality)
But the world has changed
We live in a data world
... and Black Box data is no longer fit for purpose

no provenance = no context
proprietary IDs = no connections (+ lock-in)
opaque model = limited utility
limited access = poor feedback loops

opencorporates
Today we need White Box data
... and **White Box** data is no longer fit for purpose

- no provenance = no context
- private IDs = no connections (=> lock-in)
- proprietary model = no designated utility
- limited access = no feedback loops

opcorporates
Real problem is business model of black box data providers.

That should be their problem, not yours!
Don’t be a black-box addict: the longer you stick with black-box data, the more difficult it is to change
and it’s a long road back
<table>
<thead>
<tr>
<th>Company Name</th>
<th>Location</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conifer Ridge Limited</td>
<td>United Kingdom</td>
<td>Active</td>
</tr>
<tr>
<td>Dofair Overseas Limited</td>
<td>Hong Kong</td>
<td>Active</td>
</tr>
<tr>
<td>Mossack Fonseca &amp; Associates S.A.</td>
<td>Panama</td>
<td>Active</td>
</tr>
<tr>
<td>Mossack Fonseca &amp; Co (BVI) LTD</td>
<td>Australia</td>
<td>Active</td>
</tr>
<tr>
<td>Mossack Fonseca &amp; Co (Canada) Limited</td>
<td>United Kingdom</td>
<td>Active</td>
</tr>
<tr>
<td>Mossack Fonseca &amp; Co (Moldova) Limited</td>
<td>United Kingdom</td>
<td>Active</td>
</tr>
<tr>
<td>Mossack Fonseca &amp; Co (Thailand) Ltd</td>
<td>Thailand</td>
<td>Active</td>
</tr>
<tr>
<td>MOSSACK FONSECA &amp; CO (TOLEDO) LIMITED</td>
<td>United Kingdom</td>
<td>Active</td>
</tr>
<tr>
<td>MOSSACK FONSECA &amp; CO (Ukraine) Limited</td>
<td>United Kingdom</td>
<td>Active</td>
</tr>
<tr>
<td>MOSSACK FONSECA &amp; CO ECUADOR S.A</td>
<td>Panama</td>
<td>Active</td>
</tr>
<tr>
<td>MOSSACK FONSECA &amp; CO Zurich SA (SUCCURSALE LUGANO)</td>
<td>Switzerland, 20 Apr 2016</td>
<td>Active</td>
</tr>
<tr>
<td>MOSSACK FONSECA &amp; CO (Asia) Limited</td>
<td>Hong Kong</td>
<td>Active</td>
</tr>
<tr>
<td>MOSSACK FONSECA &amp; CO (Bahamas) Limited</td>
<td>Bahamas</td>
<td>Active</td>
</tr>
<tr>
<td>MOSSACK FONSECA &amp; CO (BRAZIL) S/C LTDA</td>
<td>Brazil</td>
<td>Active</td>
</tr>
<tr>
<td>MOSSACK FONSECA &amp; CO (F.I.) Limited</td>
<td>Hong Kong</td>
<td>Active</td>
</tr>
<tr>
<td>MOSSACK FONSECA &amp; CO (HONG KONG) INCORPORATED</td>
<td>Panama, 20 Jan 1997</td>
<td>Active</td>
</tr>
</tbody>
</table>

A concrete example:

**company data**

openCorporates
the core data here is official & public ...for a very good reason
A company is not a person, but an artificial legal construct

It is essential to give “the greatest publicity to the affairs of such companies, that everyone may know on what grounds he is dealing”

– Robert Lowe, 1856
The traditional Black Box product:
- Low quality
- High latency
- Opaque models
- Proprietary IDs, classifications
- No provenance
The result

- Huge AML failures
- Bank databases not fit for purpose
- Broken government procurement
- Benign environment for criminals
- Barrier to innovation
The White Box solution

- Well-defined model
- Provenance
- Open identifiers
- Free for all via the web (search 169 million companies without registration or restriction)
Use cases:
- Foundational data
- Investigations
- Due Diligence
- Enrich/enhance

Clients (bulk data)
- Mastercard, Capital One
- Factset, Transferwise
- Law enforcement/Tax authorities
- PWC/Deloitte etc
How to break the black box addiction

1. Don’t get addicted in the first place (avoid using in new projects)
2. Don’t get in deeper (don’t add more black box data)
3. Replace step-by-step as you can (cold turkey is hard) & recognise the data-quality benefits at each stage
OpenCorporates.com

We’re hiring!