

# 20<sup>th</sup> XBRL EUROPE DAY In Copenhagen

FEBRUARY 1<sup>st</sup> 2018

*Hosted by Deloitte*

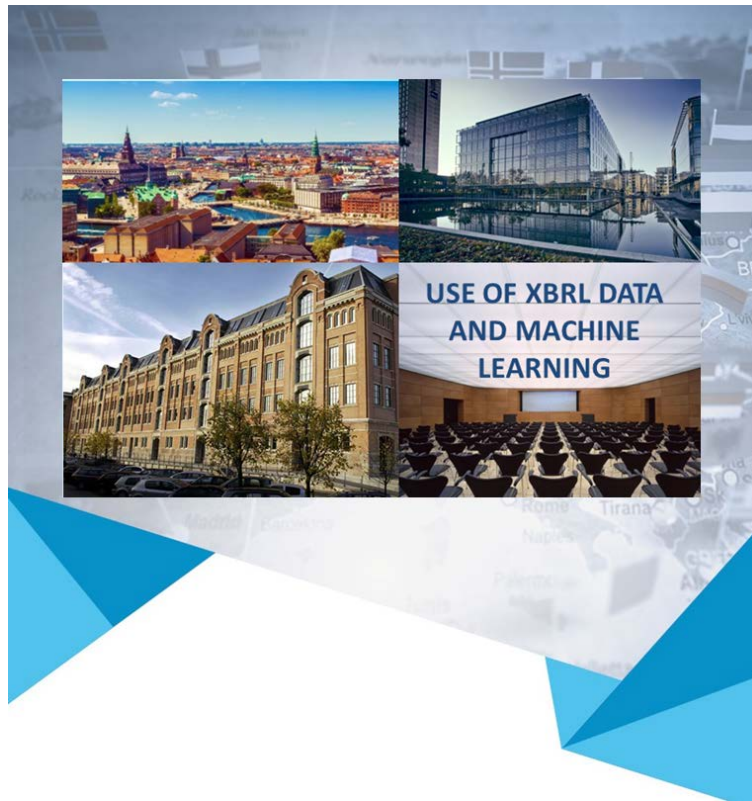
Technical Groups on Jan. 31  
*Hosted by Danish Business Authority*

**XBRL | EUROPE**

**XBRL | DENMARK**

## Structuring and restructuring OIM and Machine Learning

- Mark Goodhand
- Head of Research, CoreFiling, UK
- Chairman of XII Base Spec WG



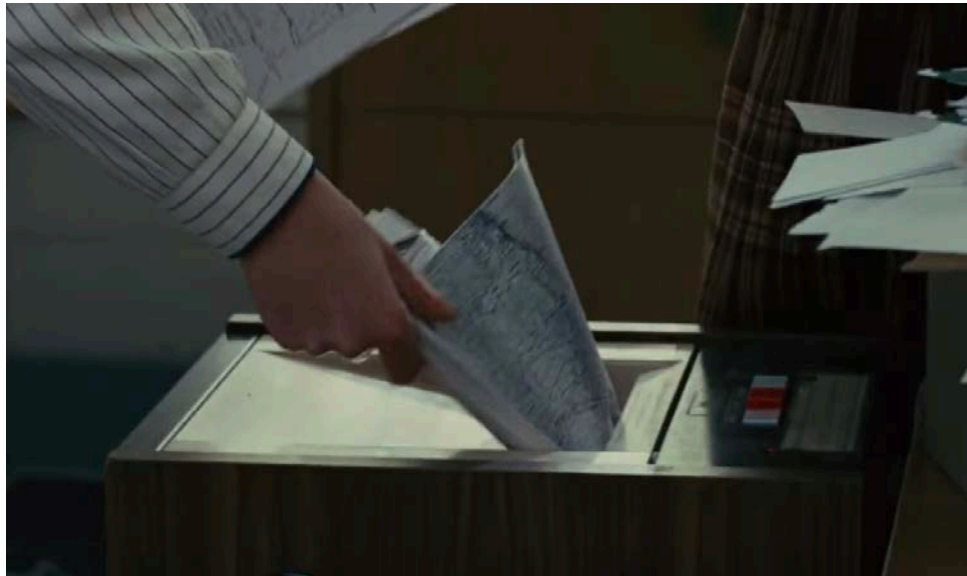
# Structuring:

Generating new insights about the world

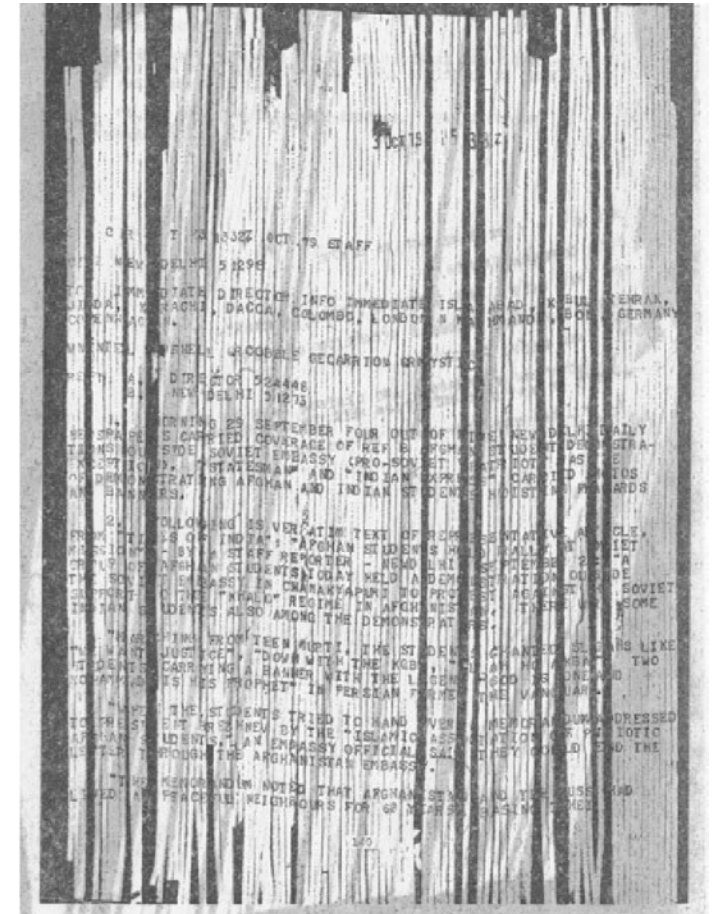
# Restructuring:

Recovering insights previously known

# Restructuring



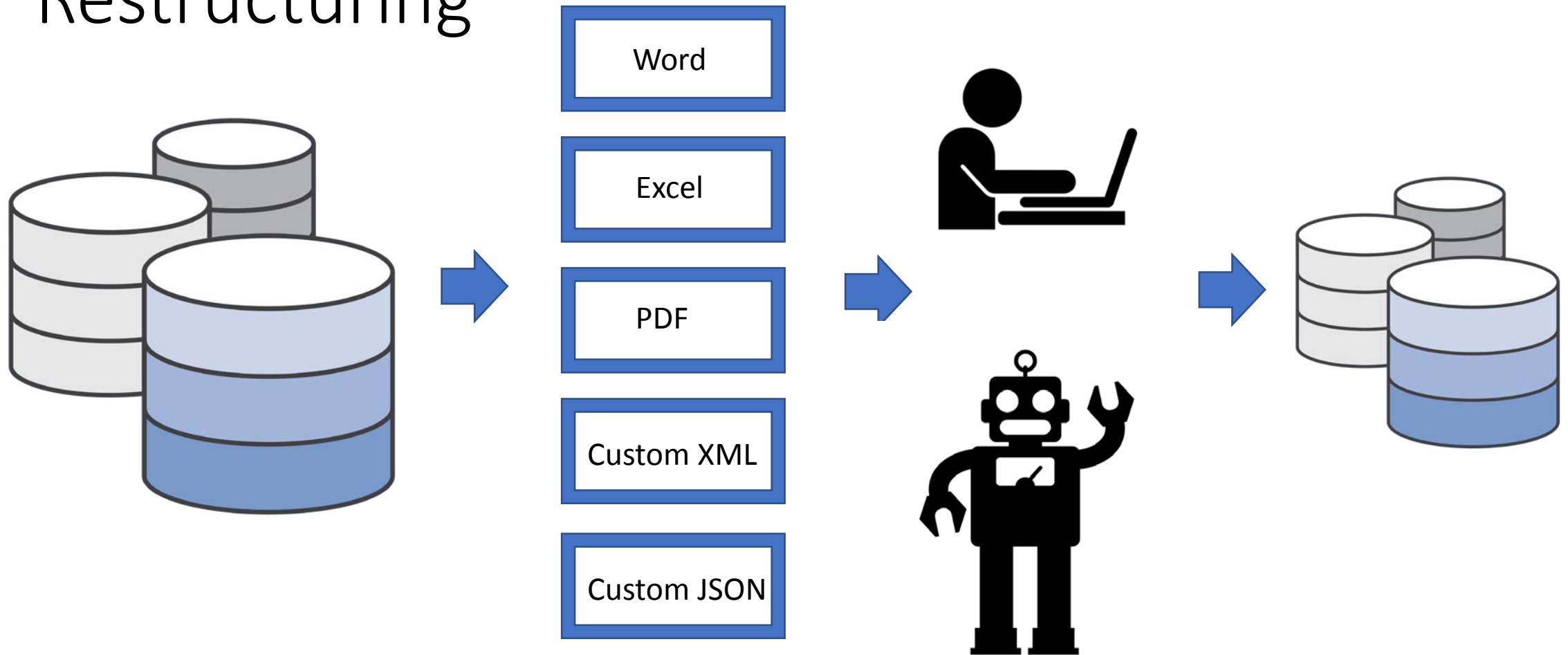
...



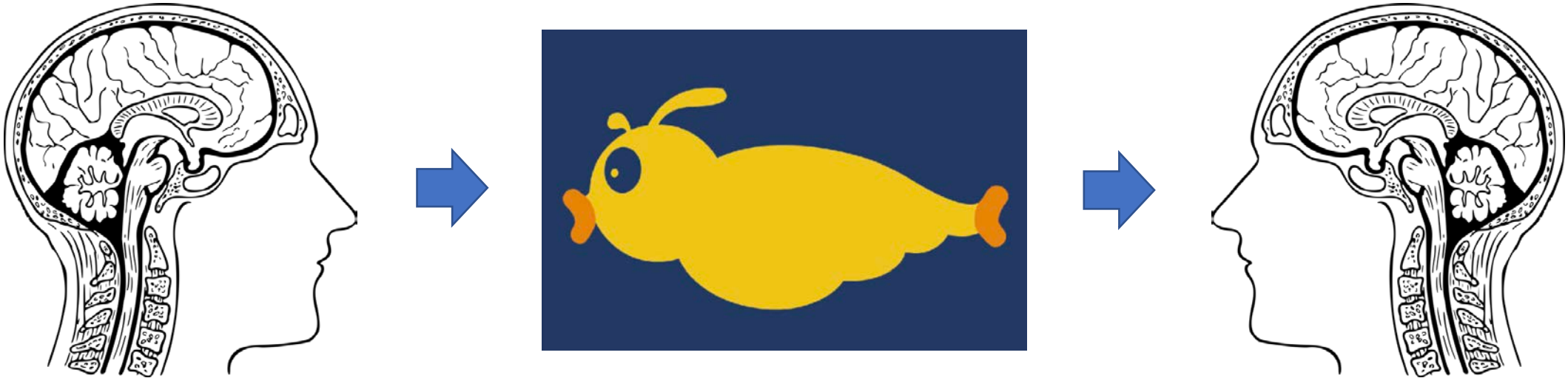
Further reading:

*Wired*: [How the CIA Used a Fake Sci-Fi Flick to Rescue Americans From Tehran](#)

# Restructuring



# Restructuring



Further reading:

- *H2G2*
- Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation

# Data Wrangling

- “Data scientists, according to interviews and expert estimates, spend from **50 percent to 80 percent** of their time mired in this more mundane labor of **collecting and preparing** unruly digital data, before it can be explored for useful nuggets.”
  - Source: [New York Times](#)

# Structuring

- “This potato has gone bad”
- “This company is about to go bust”
- “There is something unusual about this transaction”
- “You are about to be eaten by a tiger”



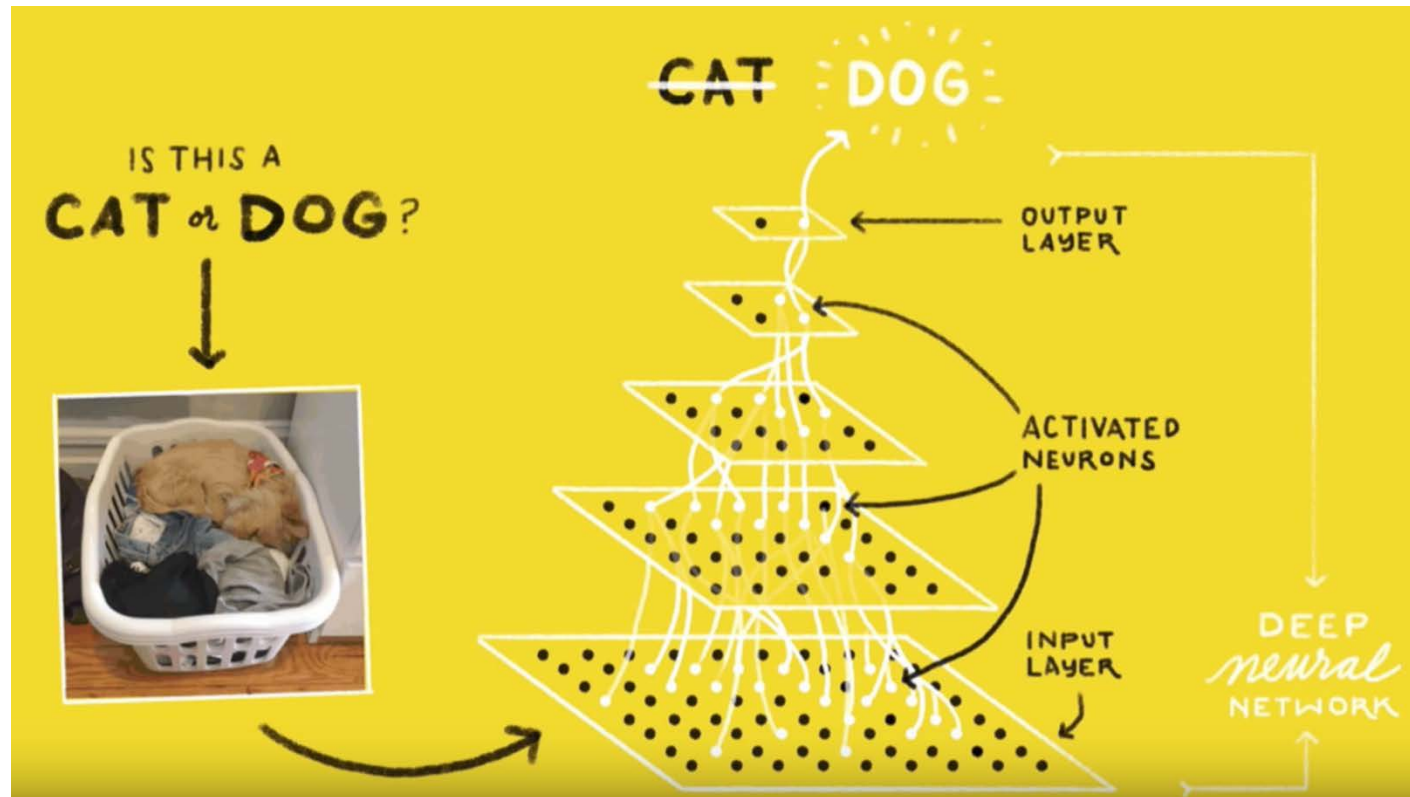
# Structuring and restructuring through Machine Learning

- Classification
- Prediction
- Clustering
- Outlier detection
- Inference engines
- Speech recognition
- Optical character recognition
- Natural language processing
- Landmark detection
- ...

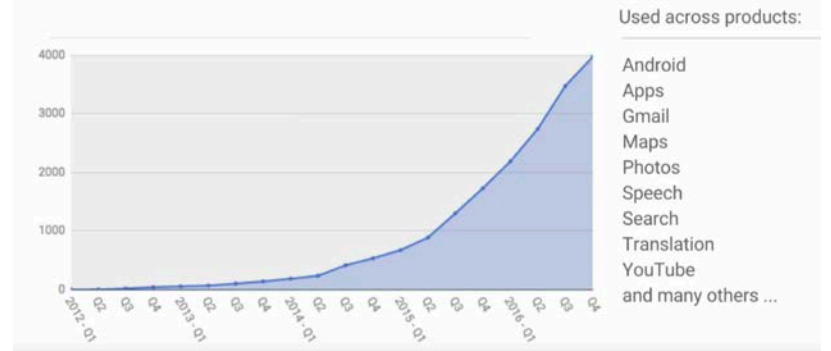


# Deep Learning

“Instead of programming a computer, you teach a computer to learn something and it does what you want.” - Eric Schmidt



Deep learning usage at Google



Source: [Infuse Your Business with Machine Learning](#)

# Automation is a boon to humanity ...

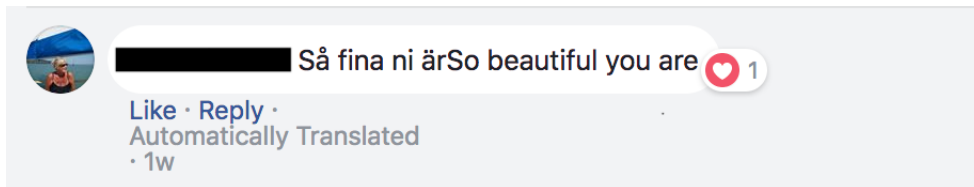
- Millions of man hours saved
- Humans saved from boring, error-prone tasks
- Barriers to communication and collaboration are reduced

... but it has its limits

- Heuristics can fail
- Deep Learning models can fail too
- Human ***judgement*** is sometimes required

# Airbnb & Facebook

- New listings & comments all the time
- Translation improves interactions
- Translations don't need to be perfect



ENTIRE APARTMENT

## Centralt placeret UNIK lejlighed

Frederiksberg



Karen Maj

6 guests 4 bedrooms 6 beds 1 bath

Translate this description to English

Lejligheden er til jer som holder af kunst og æstetik. Store smukke lyse rum med personlighed og god mulighed for at føle sig hjemme. Centralt placeret. Kun få meter til metro, handlemuligheder og Gl. Kongevej, som har skøn shopping og cafeliv. Skriv hvis I har brug for at være flere end 6 Personer.

Show original description powered by Google™

This description was automatically translated from Danish using Google Translate and may not be completely accurate.

The apartment is for you who loves art and aesthetics. Large beautiful bright room with personality and good opportunity to feel at home. Centrally located. Just a few meters to metro, shopping and Gl. Kongevej, which has great shopping and café life. Write if you need to be more than 6 persons.

# Company accounts

- Investors and regulators demand accurate figures
- Financial and legal penalties for bad data
- Human review and sign-off is essential

**SEAHORSE** The Lost Word in XBRL Filing

CoreFiling

user@example.com (Example Service Provider) | Account | Help | Logout

Home > Filers > Filings > Users

**BALANCE SHEET**  
**31 December 2010**

	Notes	2010 £	2009 £
<b>CURRENT ASSETS</b>			
Debtors	5	39,896	69,896
Cash at bank and in hand		44,738	84,738
		155,632	154,634
<b>CREDITORS: Amounts falling due within one year</b>			
	6	112,121	(119,126)
<b>NET CURRENT ASSETS</b>			
		31,508	35,508
<b>NET ASSETS</b>			
		5,506	35,508
<b>CAPITAL AND RESERVES</b>			
Called up share capital	7	200	500
Share premium	8	(2,900)	293,900
Profit and loss account	9	358,812	(258,892)
<b>SHAREHOLDER'S FUNDS</b>			
	10	13,208	35,508

These accounts have been prepared in accordance with the special provisions of Part VII of the Companies Act 1985 applicable to small companies.

Approved and authorised for issue by the board on 15 October 2011

7 / 13

**Review** << 26 of 98 >>

**Suggestions**

- 97.1% [MTS] Current assets
- 2.6% [MTS] Net assets (liabilities), including pension asset (liability)
- 0.1% [MTS] Debtors

Browse Concepts

**Concept**

Label [MTS] Current assets  
Type Instant / Monetary / Debit

**Tag Details**

Date ☐ Start of Period ☒ End of Period  
Sign As displayed

**Dimensions**

Dimension Set Operating activities, consolidated


Operating activities  
Total for all operating activities [default]

Consolidation  
Company [default]


Restatements  
Original amount [default]

Confirm Clear Cancel

# Form-based websites


 **GOING**

**MADRID (ALL) TO BARCELONA (ALL) →**


Date:  

DD / MM / YYYY

Departure ↑	Arrival	Duration	Train	Price from	Class	Rate	Options
12.30	15.30	3 h.	<span>bird</span>	<span>€ 66.75 ▶</span>	Tourist	Promo	<span>+</span>
12.40	18.10	5 h. 30 min.	AVE-LD	<span>€ 65.15 ▶</span>	<span>Tourist with link</span>	Promo	<span>+</span>

 **IDA**

**MADRID (TODAS) A BARCELONA (TODAS) →**

Fecha:  

DD/MM/AAAA

Salida ↑	Llegada	Duración	Tren	Precio Desde	Clase	Tarifa	Opciones
12.30	15.30	3 h.	AVE	<span>66,75 € ▶</span>	Turista	Promo	<span>+</span>
12.40	18.10	5 h. 30 min.	AVE-LD	<span>65,15 € ▶</span>	Turista con enlace	Promo	<span>+</span>
13.25	16.24	2 h. 59 min.	AVE-TGV	<span>107,70 € ▶</span>	Turista	Flexible	<span>+</span>
14.00	16.30	2 h. 30 min.	AVE	<span>100,40 € ▶</span>	Turista	Promo	<span>+</span>

- Relatively simple
- Slowly changing
- Can be model-driven
- Better translations by humans may be appropriate

# Once you have *structure* don't lose it!

Who, EventType, Date  
mrg, DeathByTiger, 2018-02-01

- Use open, internationally-recognised standards
- Share the information as freely as possible
- Preserve metadata
- Prefer text

```
{  
  "aspects": {  
    "xbrl:periodStart": "2018-02-01T00:00:00",  
    "xbrl:periodEnd": "2018-02-02T00:00:00",  
    "xbrl:concept": "cfl:Event",  
    "xbrl:entity": "cfl:mrg"  
  },  
  "value": "cfl:DeathByTiger"  
}
```

Enjoy the power  
and beauty of  
~~your youth~~

~~XBRL~~

OIM





# XBRL – The Good Parts

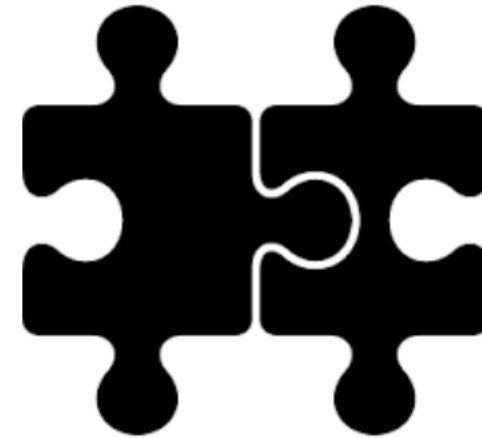
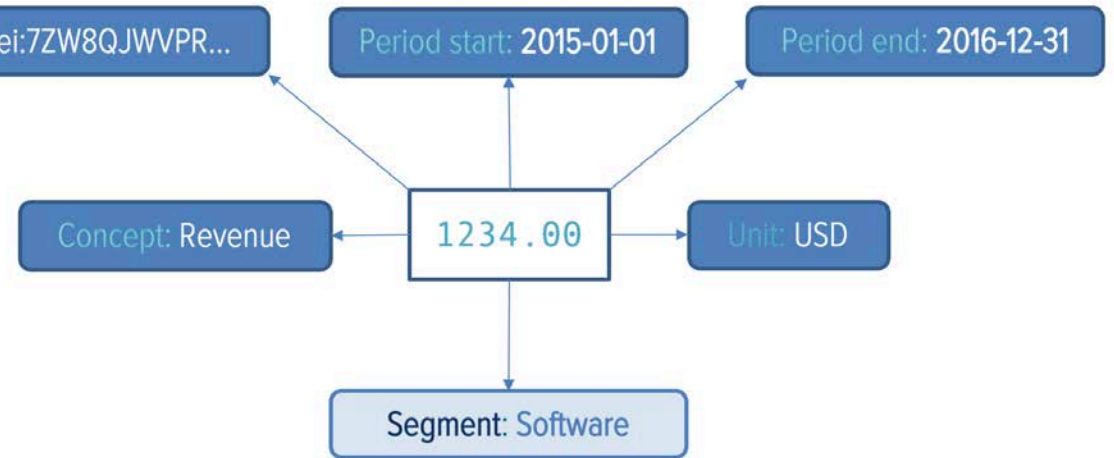
“In **XBRL**, there is a beautiful, elegant, highly expressive language that is buried under a steaming pile of good intentions and blunders.

The best nature of **XBRL** is so effectively hidden that for many years the prevailing opinion of **XBRL** was that it was an unsightly, incompetent toy.

Our intention with **OIM** is to expose the goodness in **XBRL**”

# The Good Parts

- Extensible dimensional model
- Strong types & validation
- Standards-based rendering
- Multi-language support
- Model-centric applications



# The Bad Parts

- Complex typed domains
- Segment/scenario
- contextRef & unitRef
- Custom attributes
- XLink
- Tuples

```
<xbrli:unit id='pure' >
  <xbrli:measure>xbrli:pure</xbrli:measure>
</xbrli:unit>

<xbrli:context id='CountriesHypercube_FY_31_07_2012_Set1' >
  <xbrli:entity>
    <xbrli:identifier scheme='http://www.companieshouse.gov.uk/'
    >04196519</xbrli:identifier>
    <xbrli:segment>
      <explicitMember
        dimension='ns12:CountriesDimension'
        xmlns='http://xbrl.org/2006/xbrldi'
        >ns12:England</explicitMember>
      </xbrli:segment>
    </xbrli:entity>
    <xbrli:period>
      <xbrli:startDate>2011-08-01</xbrli:startDate>
      <xbrli:endDate>2012-07-31</xbrli:endDate>
    </xbrli:period>
  </xbrli:context>
```

# OIM to the rescue

- Cut out the cruft
- Focus on the semantics
- Improve consistency
- Improve efficiency

```
{ "id": "f923",  
  "value": "1234",  
  "aspects": {  
    "xbrl:concept": "tax:NumericConcept",  
    "xbrl:entity": "cid:123456789",  
    "xbrl:periodStart": "2015-01-01T00:00:00",  
    "xbrl:periodEnd": "2016-01-01T00:00:00",  
    "xbrl:unit": "iso4217:GBP",  
    "tax:RegionDimension": "tax:Europe"  
  }  
}
```

# OIM JSON

- Clearest, simplest expression of the model
- Streaming-friendly (mostly)
- Designed for easy access to relevant information

# OIM CSV

- Efficient for large volumes of data
- Built on the W3C's CSVW (Tabular Metadata) spec
- Debate over how far we go to cope with variation in input

# The world is full of *Dimensional Data*



- Who
- What
- Where
- When
- ... Why?

```
<!ATTLIST InstantaneousBeatsPerMinute
  bpm CDATA #REQUIRED
  time CDATA #REQUIRED
>
]>
<HealthData locale="en_GB">
  <ExportDate value="2018-01-31 18:22:56 +0100"/>
  <Me HKCharacteristicTypeIdentifierDateOfBirth="1978-09-30" HKCharacteristicTypeIdentifierBiologicalSex="HKB
iologicalSexMale" HKCharacteristicTypeIdentifierBloodType="HKBloodTypeNotSet" HKCharacteristicTypeIdentifier
FitzpatrickSkinType="HKFitzpatrickSkinTypeNotSet"/>
  <Record type="HKQuantityTypeIdentifierBodyMass" sourceName="Mark's Apple Watch" unit="kg" creationDate="201
5-06-04 10:00:43 +0100" startDate="2015-06-04 10:00:43 +0100" endDate="2015-06-04 10:00:43 +0100" value="90.
4"/>
```

Expressed in a needless variety of formats



# Share and enjoy

