



# POBUCA

Your Growth. Our responsibility.

## Harnessing unstructured data

**Stavros  
Langousis**

**CTO**



# Types of data

## Structured data

- Disciplined
- Predictable
- Repeatable
- Rows and columns
- What is happening?

## Semi-structured

- Loosely organized
- Metadata provide structure
- Hashtags and folders help

## Unstructured

- No predefined data models
- Difficult to search
- Need cleaning
- Why it happened?

# Examples of unstructured data



## Text

- Documents
- Emails
- Chats
- Conversation transcripts
- Reviews
- Free text survey
- Forums



## Audio

- Phone calls
- Video audio
- IVR call centers
- Audio bots



## Image

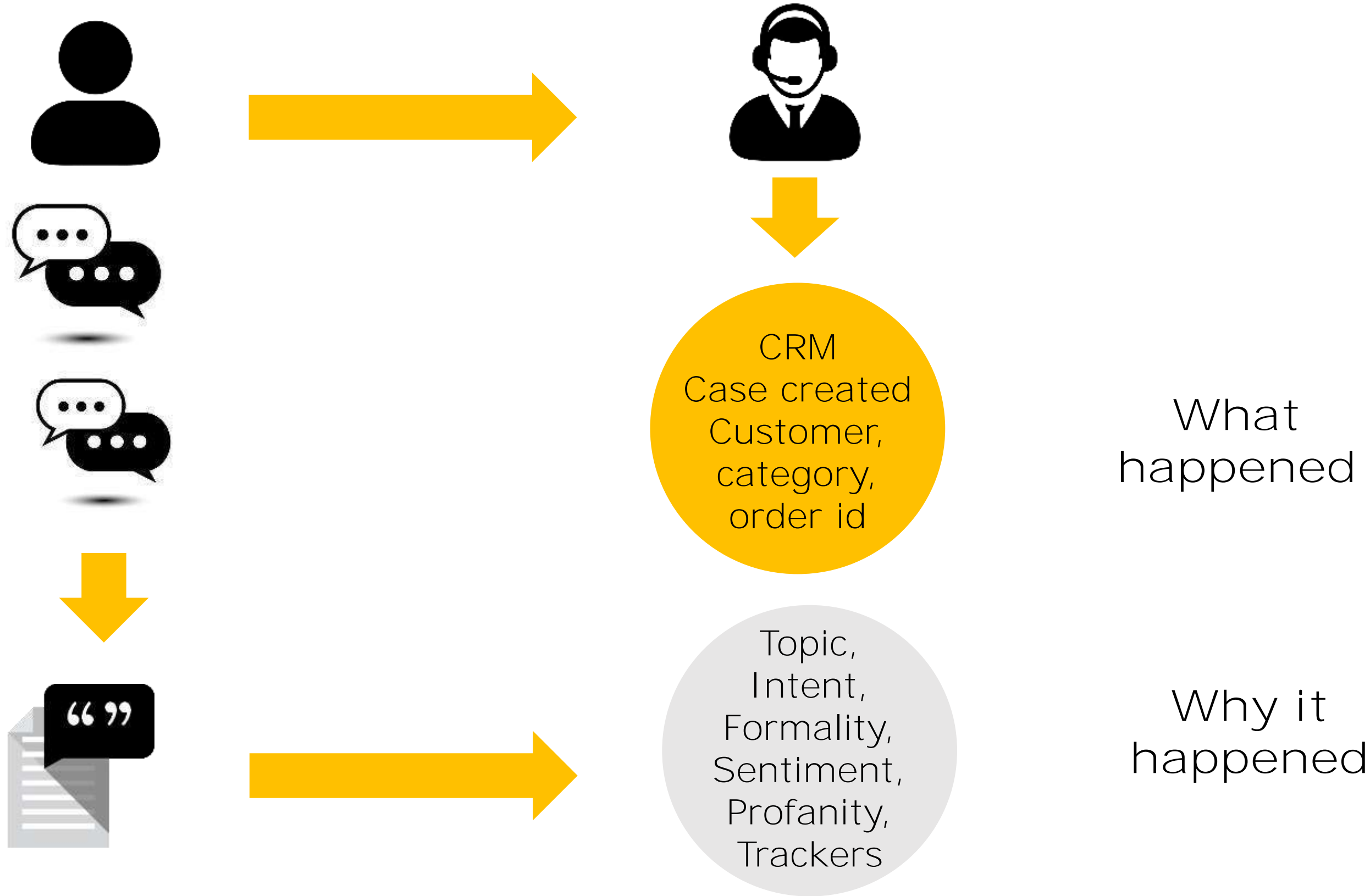
- Photos
- Images in documents
- Product photos
- Attachments
- Social media



## Video

- Surveillance cameras
- YouTube
- Video calls

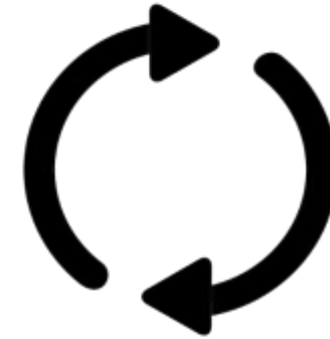
# Why are unstructured data important?



# Benefits of unlocking unstructured data



Structured data holds revenue performance figures and operational metrics, yet unstructured data can show deeper insights such as customer feedback



Feedback loop is cut short



According to projections from analysts IDC, 80 percent of worldwide data will be unstructured by 2025



Provide multi-dimensional analysis. Example: Product, customer service, shopping, brand level, last-mile



You already have them!

# How to unlock unstructured data

- 1 Data integration
- 2 Data preprocessing
- 3 Conversion to structured
- 4 Integrate with other structured data

# How to handle text data

Natural language processing (NLP) is the ability of a computer program to understand human language as it is spoken and written. It is a component of artificial intelligence (AI).



# Text annotation

I bought xyz yesterday in your shop near the airport and it broke just after 10 minutes of usage without any **apparent reason! I've** called your call center and they said they would call me back after a few **minutes but I'm still** waiting, my number is 210-2222222

Sentiment: Negative

Sentiment annotation

Product: Chair xyz

Named entity recognition

Category: Product defect

Text classification

Sub-category: It broke

Text classification

Service level: Low

Text classification

Service problem: CS Delay

Text classification

Store: Airport outlet

Named entity recognition

Date: 12/9/2021

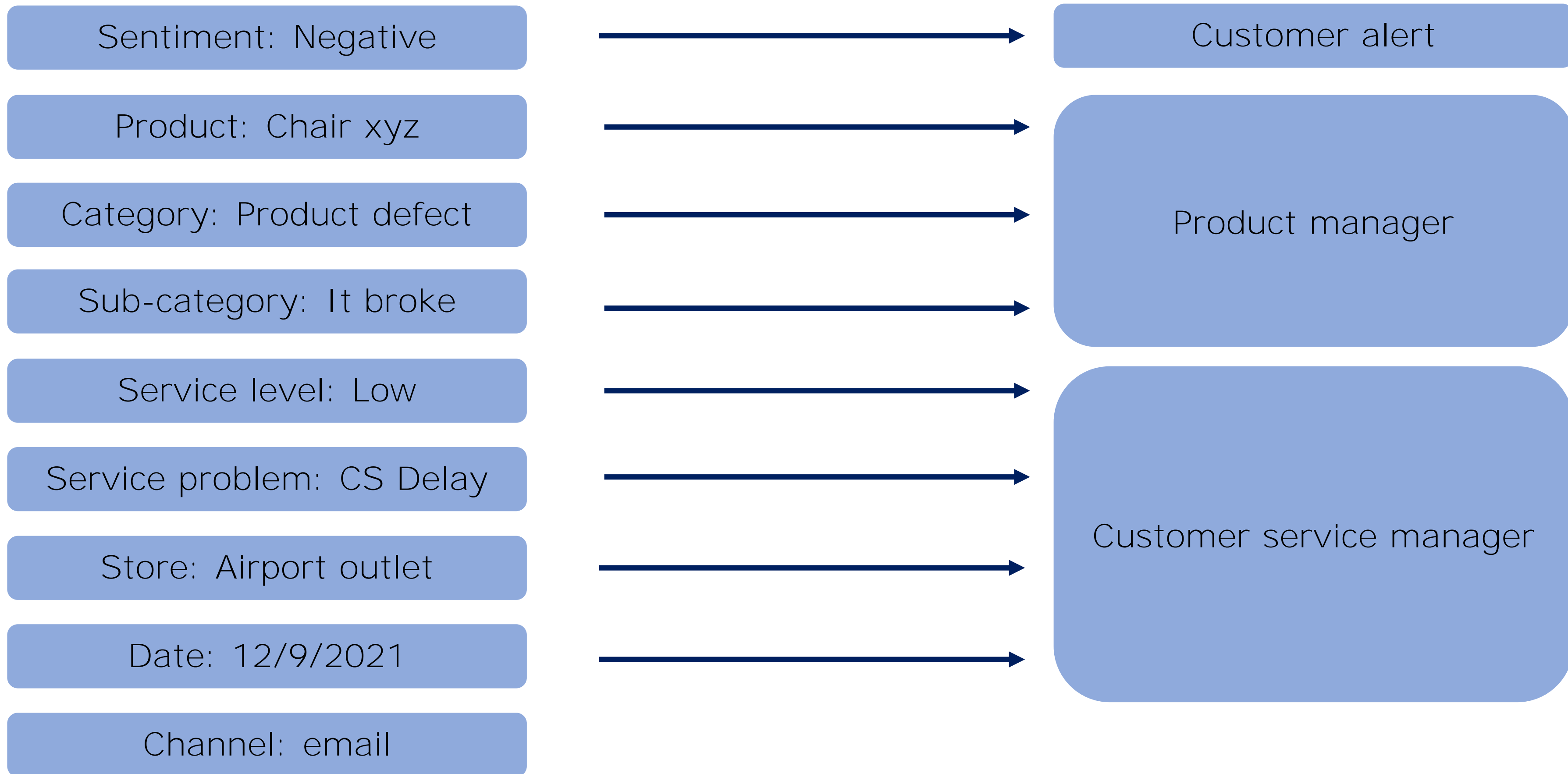
Named entity recognition

Channel: email

Metatata



# It's all about the customer experience!



Thank you!