



# AI to Capture Structured Data from Technical Manifest Documents

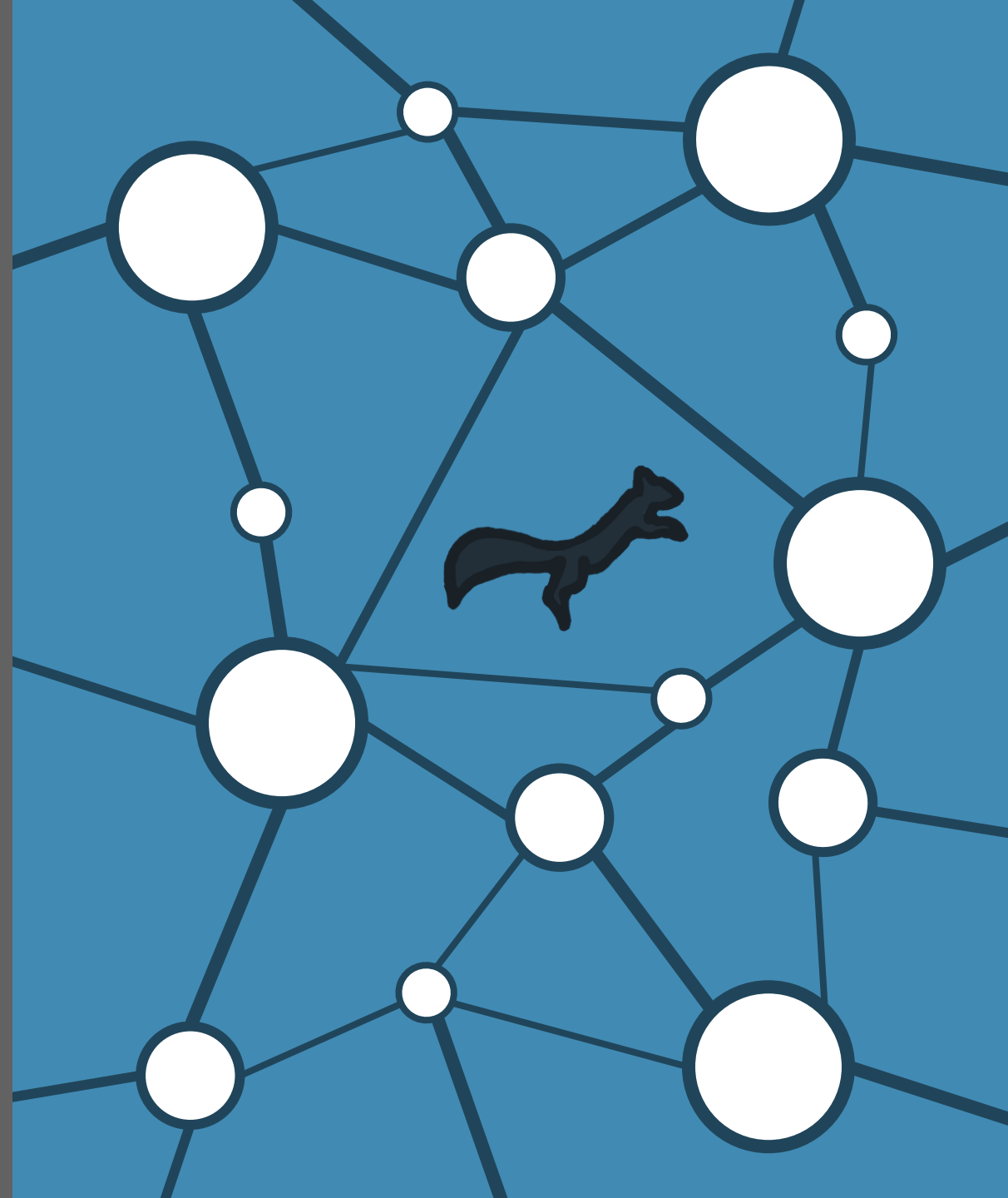
July 25<sup>th</sup>, 2023

*Amit Juneja, PhD*

*Agile Data Decisions, Inc.*

*Amit.Juneja@AgileDD.com*

*Phone: +1-617-792-5347 (M)*



- Space
  - **Intelligent Document Processing (IDP) for Manifest Analysis**
- Problem
  - **Quickly extracting useful information from cargo manifest documents**
- Solution
  - **AgileDD: a commercial human-in-the-loop AI platform that captures technical data from heterogeneous documents to automatically populate databases and automate workflows**
- Results
  - **>12X speed up in decision making from unstructured documents**
- TRL
  - **TRL 9: Used commercially in energy and mining industry**

# Human-in-the-loop data capture



Document Classification

Page Classification

Paragraph detection and classification

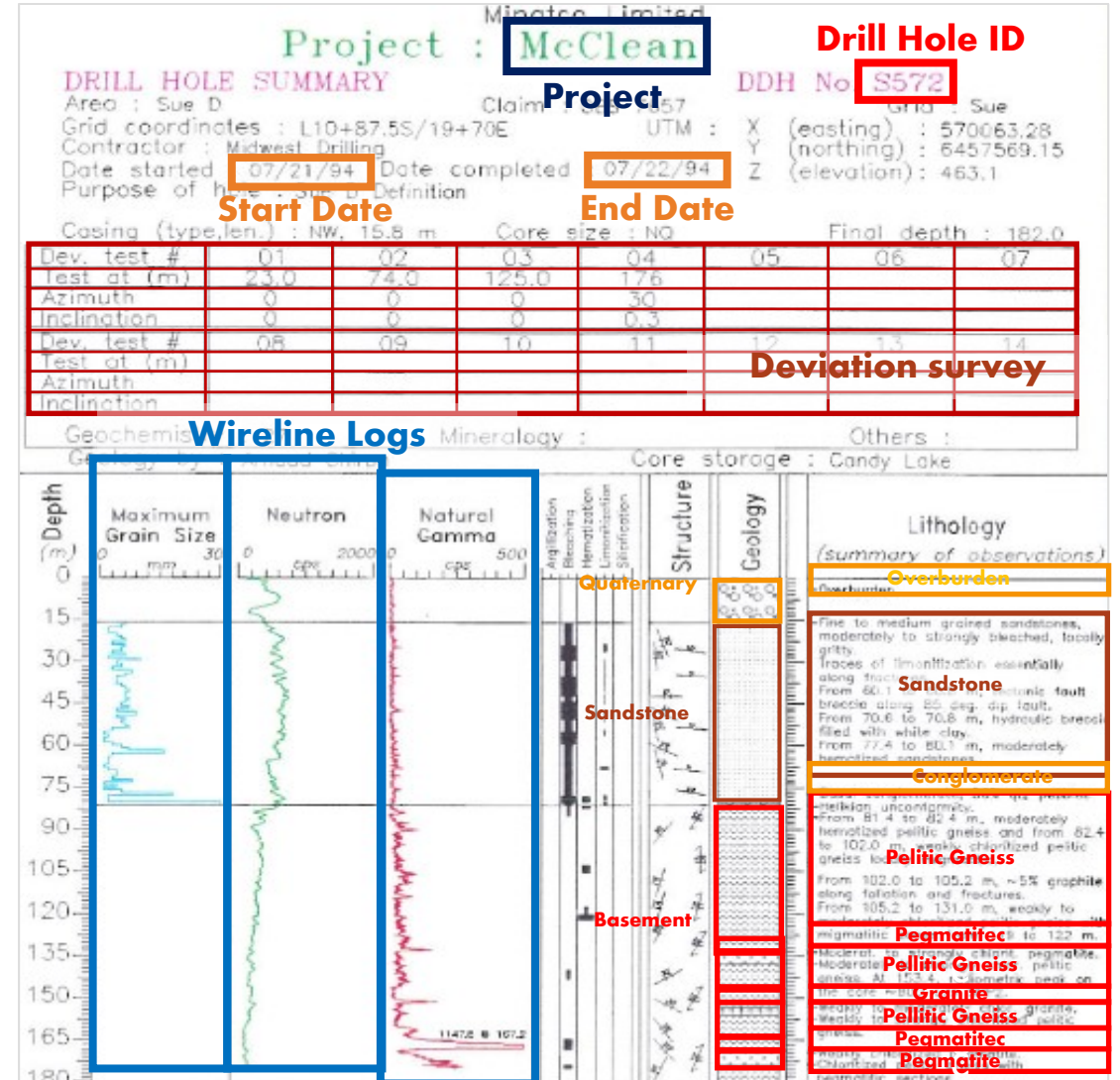
Semantic Sentence Identification

Table segmentation and classification

Visual Object Detection

Textual Information Extraction

Deployable on-premise or on cloud.  
Currently deployed in production for mining  
and energy customers. AgileDD be trained  
with CUI, proprietary or public domain  
documents depending on deployment type.



# Human-in-the-loop data capture



## Data Capture from Manifest Documents

Document Category  
Form 2134 - Shipping Manifest

# L. Shipping Manifest

**BILL OF LADING**

SHIP FROM  
NYKO Technologies Inc.  
6810 Bickmore Avenue  
Chino, CA 91708  
909-517-2200  
tony@mfa.com

SHIP TO  
GAMEPLANET INC  
7125 COFFEPOINT RD., SUITE 105  
BROWNSVILLE, TX 78521

THIRD PARTY FREIGHT CHARGES BILL TO  
NYKO Technologies, Inc.  
1990 Westwood Blvd.  
3rd Floor  
Los Angeles, CA 90025

SPECIAL INSTRUCTIONS:

Customer Order Information

Customer Order Number	# PKGS	Weight	PALLETS/SLIP	Additional Shipper Info
102808 B	36	878.4		
102808 B	24	480.0		
<b>GRAND TOTAL</b>	<b>60</b>	<b>1318.4</b>		

Carrier Information

Handling Unit	Package	Weight	H.M.	Commodity Description	LTL ONLY
3	Pit	36	Ctn	878.4	84260-05 125
1	Pit	24	Ctn	480.0	84260-05 125

Shipment Details

Ship to: Gameplanet inc 7125 coffeepoint rd, suite 105 brownsville, tx 78521  
PO1  
Cart: ST  
Title: Page 9 Confidence: [Progress Bar]

Freight Terms: Prepaid: Yes Collect: No 3rd Party: Yes

Shipper Signature: [Signature] valid Page 9 Confidence: [Progress Bar]

**Packing List**

NYKO Technologies Inc.  
6810 Bickmore Avenue  
Chino, CA 91708  
909-517-2200  
tony@mfa.com

Shipping Address  
GAMEPLANET INC  
7125 COFFEPOINT RD., SUITE 105  
BROWNSVILLE, TX 78521

P.O.NO.	FREIGHT TERMS	DUE DATE	CANCEL DATE	SHIP DATE	SHIP VIA	FOB
102808 B	FF		2008-11-18	2008-11-18	NYKO (FXNL) FedEx National	CHNO

SKU	QTY.	CONF.	SKU	RETAILER SKU	DESCRIPTION	DEST Store #	Weight
8	720	36	89018	743840000739036	Intercooler for Xbox 360		878.4
8	480	24	83060	743840000739037	Intercooler TS DS V3		480.0

Special Instructions: Y

Door: NYKO Carrier: NYKO (FXNL) FedEx National Driver Name & ID: LoadPlan: 889-260185

Trailer: SEAL Sales Order Number: 1487376

**ASN Information**

NYKO Technologies Inc.  
6810 Bickmore Avenue  
Chino, CA 91708  
909-517-2200  
tony@mfa.com

Shipping Address  
GAMEPLANET INC  
7125 COFFEPOINT RD., SUITE 105  
BROWNSVILLE, TX 78521

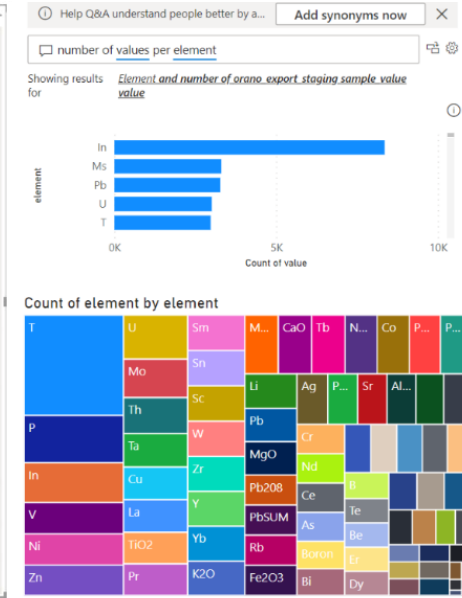
ASN date: 11/18/2008 Page 9 Confidence: [Progress Bar]

Table  
an sales number table Page 9 Confidence: [Progress Bar]

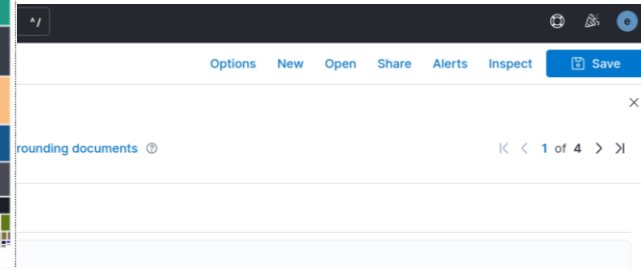
SKU	CARTONS	SERIALIZED CARTON START #	SERIALIZED CARTON END #
89018	36	0000743840000739036	0000743840000739036
83060	24	0000743840000739037	0000743840000739037

Source: Slideshare.net

# From Documents to Better Decisions



## Search and Chat



## Dashboard

Actions	Field	Value
...	@_id	rsUneyUBuxi8KaFKTF83
...	@_index	pagetext
...	#_score	-
...	@timestamp	Jan 3, 2023 @ 13:41:36.645
...	f @version	1
...	# batches_batchid	88
...	f filename	batch88/batch88/74H15-0035_ORANO_TRAINING/74H15-0035/Reports/74H15-0035_1986 Exploration, Appendices XII-XV, McArthur River Project.pdf
...	# pagenumber	54
...	f pagetext	<pre>sample: mac-148 - 498.6 field name: regolith comment: mineralogy comment 70% quartz : very angular pebbles % plagioclase : % k feldspar : % biotite .. % cordierite : % garnet : tr% sillimanite in quartzite pebbles .. % graphite .. % pyrite tr% siderite rare stringers, locally cement</pre>

## Corporate DBs



# Benefits to CBP



## Efficiency

**More quickly extract  
useful information  
from manifests and  
other documentation**

## Faster decisions

**Focus on problem  
areas &  
inconsistencies**

## Lower Risk

**Aggregate structured  
information together  
from multiple and  
hierarchical  
documents for  
analytics and “full  
picture”**



## Developing a tool for acquisition and test and evaluation analysts to:

- Capture binary and quantitative requirements from unstructured acquisition documents
- Capture test parameters, thresholds and objectives from quantitative requirements
- Automatically generate the Test and Evaluation Master Plan (TEMP) and provide links to source documents
- Reveals significant connections and discrepancies between different document versions
- Enable analysts to more easily identify patterns and significant gaps
- Will deliver prototype by FY24 Q1
- Work being performed under STTR Phase II contract with AFWERX for US Air Force Test and Evaluation

# Reach out for demo!



The screenshot displays the AgileDD application interface. At the top, there's a navigation bar with 'Workspace', 'Attributes', 'Train Models', 'Admin', 'Edit Units', and 'Project Settings'. A search bar is present on the right. The main content area shows 'Requirements128.pdf' with a document category of 'Unclassified'. Below this, there's a 'Page Category' dropdown. The central pane displays extracted text from the document, including 'Performance Work Statement (PWS) Space Control Depot Support (SCDS) II', 'Solicitation No. FAxxxx-10-x-xxxx', 'Contract No. FAxxxx-10-x-xxxx', and '10 Page of 39'. The text is structured into sections: '3.2.2.3. CM Library', '3.2.2.4. Configuration Control', and '3.2.2.5. Configuration Audits'. A large red 'DRAFT' watermark is overlaid on the text. On the left, a sidebar shows a list of requirements with filters and search options. On the right, a vertical pane shows page thumbnails for pages 10, 11, and 12, each with a 'DRAFT' watermark.

Amit Juneja, PhD  
Agile Data Decisions, Inc.  
Amit.Juneja@AgileDD.com  
Phone: +1-617-792-5347 (M)

*“AgileDD.ai is relatively unique with their patented approach for modeling the extraction process in the visual plane of a document, rather than treating a document as a sequence of words reading from top to bottom. Furthermore, any acceptable solution also has to be attuned to the semantics of the particular documents’ business domain, jargon, nuances, terminology and the user’s objective. This venture is on the right track.”*

