



# **Big Data Analytics and Machine Intelligence Strategy**

Manjula Ambur

NASA Langley Research Center

April 2014



# Outline

- What is Big Data
- Vision and Roadmap
- Key Capabilities
- Impetus for Watson Technologies
- Content Analytics Use
- Potential use cases



# What is Big Data ?





# Big Data Analytics and Machine Intelligence Strategy

**Objective:** Enable NASA employees to utilize and apply these transformational technologies as force multipliers for scientific and engineering discoveries and systems innovation and optimization

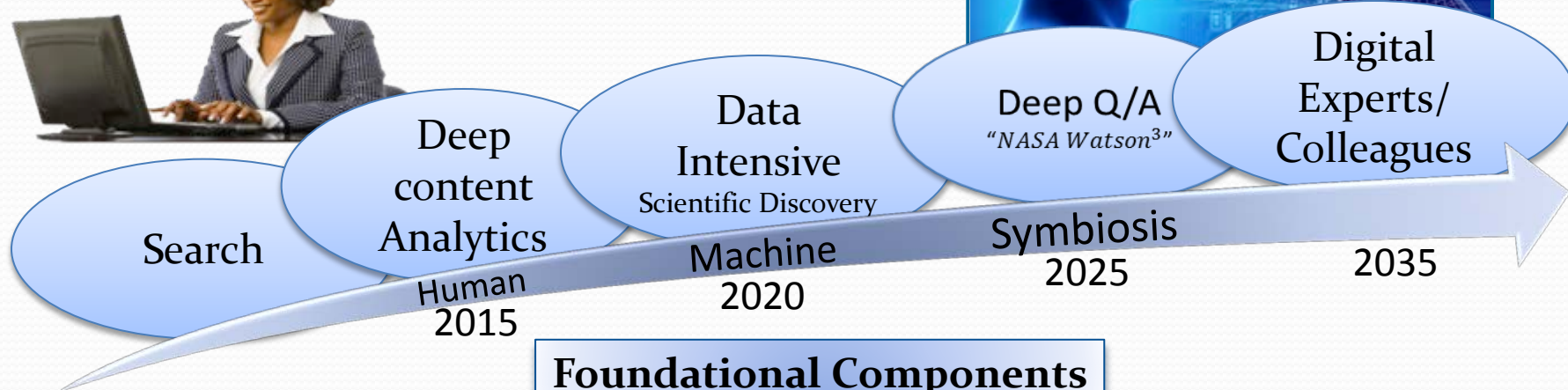
- **Vision:** Researchers, Engineers, and Project Teams have “Machine/Virtual Expert(s)/Colleague(s) and Personalized Intelligent Agents at their disposal that can:
  - Digest, synthesize, and keep up with global knowledge
  - Answer specific questions
  - Synthesize & makes sense of volumes of big and heterogeneous data/information – data intensive scientific discovery
  - Provides predictions for new technologies and design configurations
  - Processes modeling & simulation data in real time
  - Human cognition and machine cognition augmenting each other providing unimaginable capabilities
  - Eventually machine experts and human experts working side by side

**Team:** Thought Leaders; Researchers; Engineers; IT Specialists; Statistician; Computer Scientist



# Big Data Analytics & Machine Intelligence Capabilities - Roadmap

Human Experts and Digital Experts augmenting and learning from each other in an organic way



Data Mining		Machine learning Semantics		Knowledge Technologies In-Situ Analysis Algorithms		Data Visualization Ontologies		Content Analytics			
Web sites Open web		Images Documents		Knowledge Base Data sets Videos		Global Sources Databases		Distributed data stores Journals Blogs			
Digital Notebook		Data Integration Data Curation		Data Management & Collaboration Analytics Sent to Data				Data Capture Meta Data Tagging		Remote Access	





# Two Key Capabilities towards Vision

## Deep Content Analytics

Application of sophisticated natural language processing and machine learning techniques to large corpus of knowledge to obtain insights, trends, and answers to specific questions.

## Data Intensive Scientific Discovery

The 4th Paradigm – Advancing from hypothesis based experiments and mod-sim to data intensive scientific discovery; deriving new insights and correlations not possible otherwise

2012 – Content Analytics Pilots begin

2014 – Knowledge Assistant Pilot - Incubators start using analyzing 100K plus articles

2016 – KA for one discipline with trends and alerts

2017 – Knowledge Assistant for core disciplines

2020 – NASA Watson Pilot with multimedia and Q & A capability

And more...

2015 – Expand pilot program into more disciplines

2020 – Real time data analytics for Mod Sim data pilots

And more....

2014 – NDE, Aeroelasticity, & Incubator data pilots start

2018 – Establish Data Discovery Capability in a few disciplines

2020 – Automatic data capture, tagging and integration

*NASA Watson<sup>3</sup>*

Virtual Research & Design Partner



# Impetus for Watson Technologies

- Center Focus on Knowledgeability and Technical Excellence
  - Scientists and engineers access, search, find, integrate, synthesize and digest global knowledge - Beyond Search
- Started with Federated/Integrated search
  - Enterprise Google implemented in 2006 with good results
- Investigated semantic technologies
  - Found to be resource and subject matter expertise intensive
- Started to investigate text mining and data mining technologies
  - IBM Watson made a big flash and started that journey.
- **Vision: Wearable or Embedded Intelligent Agent**
  - Configurable and personalized intelligent agent that is wearable to embedded activated by voice or even brain waves



# NASA Langley Watson Journey...

## **Investigation: 2011**

- 2011: Center wide IBM Watson Seminar by IBM Expert
- 2011: Center Workshop: Developed Use cases

## **Prototype: 2012**

- Visit to Watson lab and discussions with IBM Experts
- Decided to experiment with IBM Content Analytics, a key component of IBM Watson ; cost effective starting point
- Successful Prototype with IBM Content Analytics to apply Advanced Content Analytics (ACA) techniques & methodologies to 3 use cases
  - In collaboration with mission organizations and IBM Experts

## **Pilot: 2013**

- Pilot with Advanced Content Analytics: 4 use cases
- Workshop with Senior Leaders and Researchers for next steps
  - Decided the focus to be Knowledgeability and Innovation – ‘Knowledge Assistant’

## **Capability: 2013-2014**

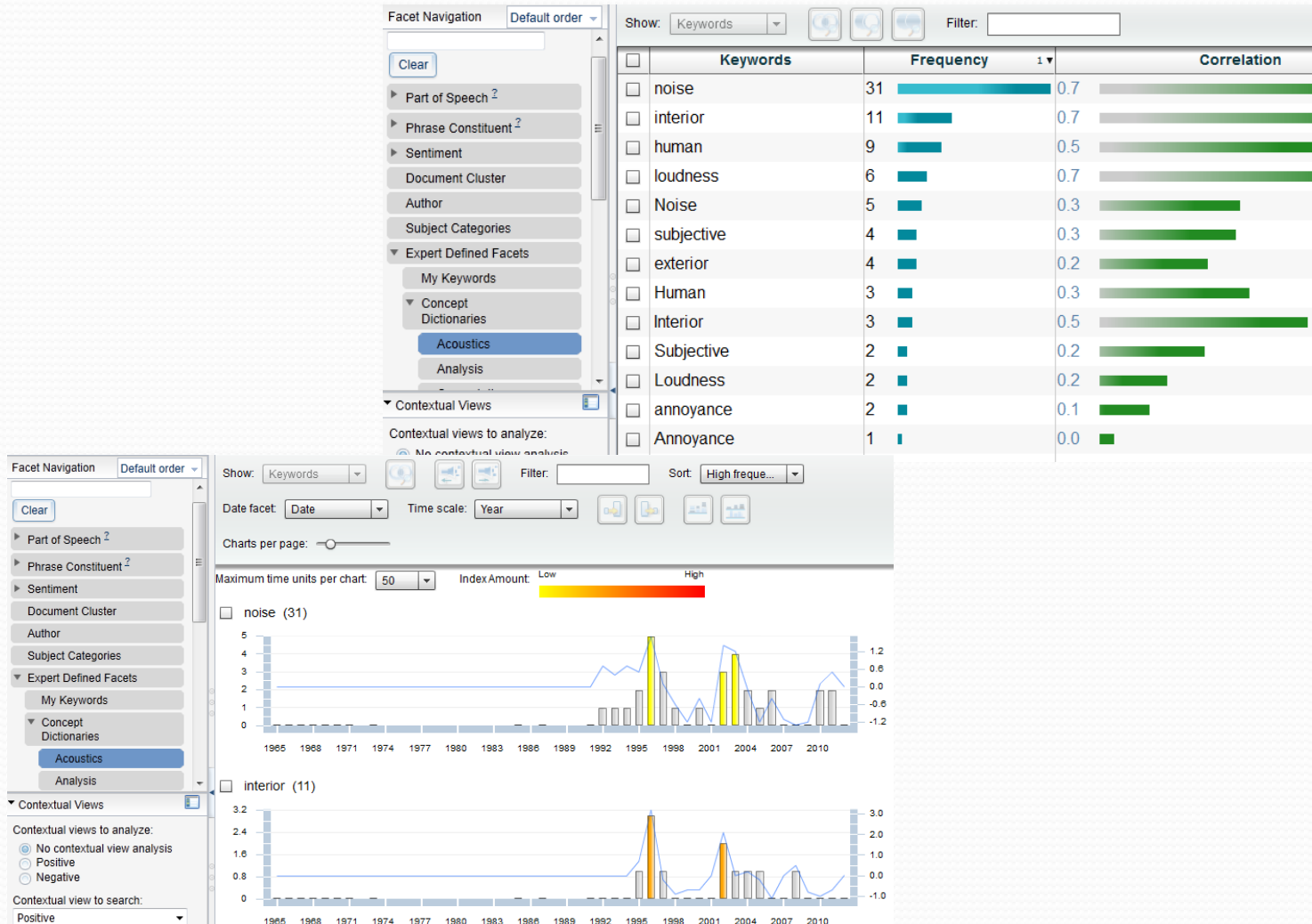
- Advanced content analytics being offered as part of OCIO capabilities/Services and as part of our ‘Mining for Knowledge’ sessions
- Knowledge Assistant Pilot being formulated in specific disciplines with Q and A capability – beginnings of “NASA Watson”





# What is Content Analytics?

*Content Analytics* refers to the **text analytics process** plus the ability to **visually identify and explore trends, patterns, and statistically relevant features** found in various types of content





# Use Cases Examples

- **Analysis of Sonic Boom Research**
  - Focusing on specific areas with out reading ; used Automated clustering and categorization techniques; ~1500 Reports
- **Analysis of National Safety Board Accident Reports**
  - Gain better Insights and save time in analysis; analyzed ~3600 reports
- **Finding Technology Opportunities from FBO.gov**
  - Analyze data looking for opportunities - trends, experts in chosen technology areas.  
Data Source: 120,000 XML current and archived records
- **Analysis of Structures and Materials Publications**
  - Concept search, pattern analysis and classification of publications in structures and materials areas from 14,300 publications/reports
- **Research Opportunities in Autonomous Flight areas**
  - Subject search and analysis, trends, experts and opportunities (current, emerging and niche) in fields related to autonomy; 1,500 documents from many different sources (NASD, AIAA, Engineering Village, etc, ...)



# Knowledge Assistant

*Creating a Research Assistant / Virtual Team Member*

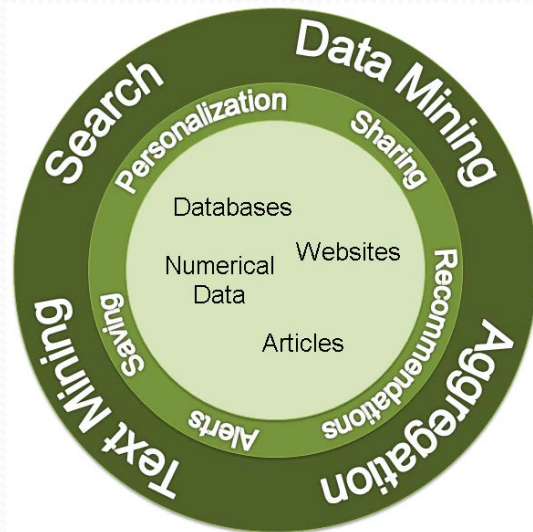
**Purpose:** Enable and Improve Center Knowledgeability and Innovation

## Current Methodology:

- A significant amount of time is spent mining for targeted knowledge, manually by SME.
- Data sits unexplored. Connections not made.
- Insights missed.
- A knowledge assistant would serve as a virtual colleague.

## Goals:

- Keeping up with technical and competitive intelligence
- Making sense quickly: Find wheat in the chaff.
- Identify Strategic business opportunities
- Enable cross Discipline Innovation
- Identify and connect networks of experts.



## Value:

- Help/Improve Center Knowledgeability - **Market/Competitive/Technical Intelligence**
- Identify key **trends**, emerging **experts** and **expert networks**; **summaries**, **alerts**, **recommendations**, non-obvious **relationships** and **intuitive visualization of results**
- Give users the ability to ask questions and get answers -- **Deep Q&A**



# Potential Use Cases

- Enable better and faster decision making utilizing unstructured big data
  - Data intensive scientific discovery – Fourth Paradigm
  - Knowledge discovery and mining
- Predictions for business/technology opportunities
  - Machine-automated survey of engineering / science trends worldwide
  - Determine emerging trends
  - Find breakthrough connections among seemingly un-related disciplines
- Analysis & visualization of Big data: PetaByte-sized and rapid-flow real time data and information streams
  - Computational Fluid Dynamics, sensor data analysis & visualization
  - Enables Simulation based science and engineering helping to reduce the time and enabling lab to computer interactions and synthesis
- Deep Q and A system using Cognitive and computational Knowledge engine
  - Answers to specific engineering and aerospace questions