Strengthening the decision making process with data intelligence in publishing industry

CONTEC 2014 – Frankfurt Germany - October 7th 2014

Vincenzo Russi
Chief Digital Officer
Messaggerie Italiane SpA
Objective

Publishing organizations are looking for new sources of guidance to support their decision making, and efficient new techniques to enable them to collect data from disparate sources and cost-effectively convert it into actionable insights. Presentation illustrates how enterprises gain business efficiencies by meeting internal and external demands for ‘information’ rather than 'data', and how this enables new interactions with all the publishing value chain stakeholders: authors and readers, publishers, distributors, wholesalers and booksellers.

Release a Group Project cross companies and cross functions (marketing, sales, IT, operation, management)
Introduce a new paradigm for digital projects
Found new rules for technology driven projects (exceeding the traditional time, budget and spec)
Break the traditional approaches and get better expectations
Prove the value of data driven decision support
Deliver a company definition for data intelligence
Develop a new data culture in the organization
# Containers of Content

<table>
<thead>
<tr>
<th></th>
<th>paper</th>
<th>ePaper</th>
<th>web</th>
<th>Smartphone</th>
<th>Tablet</th>
<th>Smart TV – Social TV</th>
</tr>
</thead>
<tbody>
<tr>
<td>reliable</td>
<td>always available</td>
<td>updated social service oriented</td>
<td>instantaneous short localized</td>
<td>interactive</td>
<td>rich</td>
<td></td>
</tr>
</tbody>
</table>
Digital Continuity

Breakfast 6:00 am
Transfer
Office
Meeting
Lunch
Office
After Dinner
Sleep 11:00 pm

Tablet
Mobile
Laptop
Tablet
Mobile
Laptop
TV Tablet
Tablet

© Messaggerie Italiane SpA, Frankfurt Book Fair, 2014
The Digital World and Data «Caos»

Source: BI Intelligence, 2014

Source: IDC, BI Intelligence Estimates

*For context: Facebook ingests 500+ TB of data each day

68% of all unstructured data in 2015 will be created by consumers (*)

90% of all data was unstructured

Source: BI Intelligence, 2014
Publishing Value Chain and Data Complexity
Messaggerie SpA Organization and Information Cluster

PUBLISHING

DISTRIBUTION

RETAIL

© Messaggerie Italiane SpA, Frankfurt Book Fair, 2014
Market Analysis

Internet Value  

Volume  

Value  

Sources: Market Analysts
The Urgency for Data Driven Decisions

**Emerging Actors**
- Authors
- Retailers
- Tech companies
- Agents

**New Publishing Models**
- Self-Publishing
- Social Publishing
- Self-Publishing
- Co-Publishing
- Pop Publishing
- Crowd Publishing
- Agile Publishing

**Stakeholders and Market**
- Authors and Readers
- Publishers, Distributors, Wholesalers, OnLine and Offline Retailers, Agencies
- Price, margin and profit turbulence
- Digital and Physical opportunity
The Traditional Method for Data Projects

- **Requirements**
- **Analysis**
- **Design**
  - Conceptual
  - Logical
  - Physical
- **Test**
- **Delivery**

*High costs due to ownership negotiation, confidentiality, durability, reliability, contractual, ...*
Answer New Questions or Unknown Questions
Speed and Iteration: Project Spiral Approach

Characteristics

Stage-based
Agile project management
Flexibility, Resiliency and Responsiveness
Quick release of incremental deliverables and integration of client feedbacks towards the next version
Tested method for innovative projects

Design Sources
Engineering Sources
NewTeam Sources

Design and Develop (Next) Version
- Problem Framing
- Mock up and Concept
- Visual Prototyping
- Prototype and Candidate Engineering

Ideate (Next) Version
- StartUp and scouting increase the project success probability and accelerate the iterations

Collect Feedbacks
Test and evaluate with stakeholders and target client persons

Plan (Next) Iteration
Each iteration increments functionalities and performance

Source: Vincenzo Russi

© Messaggerie Italiane SpA, Frankfurt Book Fair, 2014
Qualitative and Quantitative Comparison

Key Figures

- Months project duration as planned
- Multi companies involved
- Stakeholders involved in non-invasive meetings and interviews
- Operating professionals involved in definite data collection and ingestion
- Project leader team: group objectives, data, technology
- Technology and process consultants
- Cost less than the forecasted budget
- Board and Shareholders support and contribution
- Over expectation deliverables, performances and functionalities

Phases general cost in percentage of the total cost

Pie charts areas are proportional to project costs
New Paradigm for Data Driven Decision

Schema on Write model

<table>
<thead>
<tr>
<th>Table</th>
</tr>
</thead>
<tbody>
<tr>
<td>PK</td>
</tr>
<tr>
<td>attribute name</td>
</tr>
<tr>
<td>attribute name</td>
</tr>
<tr>
<td>attribute name</td>
</tr>
</tbody>
</table>

Schema on Read model
Data Lake

Transaction and Network  Desktop and Mobile  File Formats  Social  Contextual

Paid market research
Consumer behavior/behavioral
Adv/publicity/discount campaign
Marketing profiling

Raw Data Stream

Data Lake
Information Mill
Decision Making

Informed

Decision Making
Data Platform

DATA SOURCES
- Transactions, OLTP
- Documents and eMail
- Social Media, Web Logs
- Machine Device, Scientific

DATA INGESTION
- Batch Processing
- Real Time Streaming

DATA INTEGRATION

DATA LAKE

DATA ANALYSIS
- Traditional BI Apps
- Analytics
- Data Exploration

DATA GOVERNANCE: METADATA, DATA QUALITY & MDM, DATA SECURITY

Source: Reply SpA
Layers

Multiple layers approach

Data Ingestion
Layer for data import into the Big Data architecture.

Data Storage
Layer for data Persistence.
The Data Storage layer deals with data persistence inside the Big Data architecture. Specifically, HDFS, the Hadoop distributed file system, constitutes the data logging layer. HDFS, in addition to the ability to manage large data sets, is particularly efficient on large sequential reads and possesses mechanisms to manage the data consistency and replica.

Data Processing
Layer for parallel data computation on a distributed environment.

Data Analytics
Layer for data processing.

Data Visualization
Layer that enables the final users to access the analysis results.
Data Quality

Quality of data inter Information Systems
Front end analysis support for a rapid quality assessment
Graphical user interface provision for data overview
Graphical representation and flexible queries can help in perception of the quality
Data Obsolescence, Resilience, Relevance

Single data becomes obsolete by definition
Data Obsolescence, Resilience, Relevance

*Single data becomes obsolete by definition*
Data Obsolescence, Resilience, Relevance

Single data becomes obsolete by definition
Data Obsolescence, Resilience, Relevance

Single data becomes obsolete by definition
Data Obsolescence, Resilience, Relevance

An integrated data storage of multiple data streams helps to protect the data value against obsolescence, by extending its life and relevance.
Dynamic Costs

**Dynamic variables**

- Data ingestion
- Data sources
- Data volume
- Query complexity
- Analysis depth
- Users (# and skills)
- Frequency (users and queries)

Data Space

Consultancy

Services and Platforms

Computation and Processing

- On Demand based on real needs and business oriented
- Services from suppliers, Open Sources platforms
- Dynamic allocation of computational nodes and on demand provision for data space
Data Intelligence Culture

Lower the entry level barrier for data analysis

Gather requirements from stakeholders, both internal and external

Extend to the enterprise the usage of data analysis and the value perception of data

Exploit moments, meetings and events to develop the culture

Implement a «liquid» environment to adapt different users and skills

 Transmit confidence on data analysis through views and security

Accept data and information constraints

Provide a reliable environment from quality, robustness, availability and performance point of view

Consider the usability and experience as a key driver for user acceptance
Data Driven Digital Ecosystem

+ Social networks and communities systems
+ Relations, contacts and campaigns management
+ Search, visualization and identification systems
+ Digital devices exploitation
+ Digital commerce
+ Data analysis on big data
+ Localization management
+ Innovative usage of physical presence
+ Logistic and resulting distribution systems
+ Advertising models
+ Content management
+ Development and as-a-service adoption of technology
Components, Grammar, and a New Composition

Data stream in data lake are like Jabberwocky in «Through the looking glass»: unintellegible language.

Recognizing the «verse of data» using the «right mirror glass» we can read what is considered one of the greatest poem written in English that has given to the language «neologisms».

Through the looking glass. Lewis Carrol
Scene with the Red Queen and Alice. Alice is constantly running but remaining in the same spot.

“Well, in our country,” said Alice, still panting a little, “you’d generally get to somewhere else — if you run very fast for a long time, as we’ve been doing.”

“A slow sort of country!” said the Queen. “Now, here, you see, it takes all the running you can do, to keep in the same place. If you want to get somewhere else, you must run at least twice as fast as that!”

—I’m not «Alice»
— I’m not the «Red Queen»
— I have to be a «Digital Runner»