

Data Journalism in RAI

Authors

Giulia Dezi, Giorgio Dimino, Maurizio Mazzoneschi,
Alberto Messina, Sabino Metta, Giuseppe Mondelli, Maurizio Montagnuolo

RAI – Radiotelevisione Italiana

Centre for Research and Technological Innovation

RAI Digital

Presented by Dr. Alberto Messina

RAI – Centre for Research and Technological Innovation

6th EBU MDN Workshop

Geneva June 7th 2016

Agenda

- ◆ **Definitions**
- ◆ **Best practises**
- ◆ **Our approach**
- ◆ **The data team**
- ◆ **Tools & architectures**
- ◆ **Some preliminary results**
- ◆ **Development ideas**
- ◆ **Conclusions**



Role of Data in Journalism

- ✓ as origin of the news and of related content
 ✓ (*data driven journalism*)



- ✓ as accurate and verifiable description of inherent semantic aspects
 ✓ (*precision journalism*)



- ✓ Basic building block for presentation / usage of news content
 ✓ (*visual journalism*)



◆ **Extracting sense out of data to create newsworthy stories**

◆ **This implies**

- Having data
- Analysing data
- Identifying “sense”, or “sense the news”
- Presenting data

Data ➤ **Analysis** ➤ **Story** ➤ **Publication**



Data Driven Journalism some insight

Data ➤ **Analysis** ➤ Story ➤ Publication

Data Harvesting

Which data?
Which formats?
Which time window?



Data Analysis:
Content Analysis
Semantic Analysis
Statistic Analysis



Data Classification:
Ontologies
Automated categorisation

Team
Collaboration

Editorial Process

Detecting the Story
Modelling the Story
Developing the Story

Platform Adaptation
Second Screen
Interactivity





Best practises



The New York Times

BBC

theguardian



LA NACION

ICIJ

★ THE TEXAS TRIBUNE

Editorial objectives

Production & org. approach

Wanting a **100% digital product**

Huge investments in integrated tech. resources

Data Journalism as a **public service**

Visual Journalism and ad hoc apps

Continuous improvement of **product quality**

Perfectly integrated desks (journalists & techies)

Data as **source for narration**

Notable **technical competence** of journalists

Open Source and **Open Data Journalism**

Crowdsourcing as a resource

Investigative Journalism through data

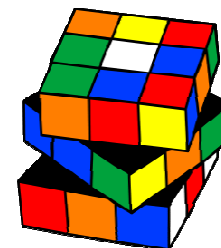
International collaborations for data exchange

Expanding scope from local to international

Experimental editorial techniques



Our Approach



- ◆ **RAI wants to foster an approach at Data Journalism orientated to**
 - Increase information **trustworthiness**, based on harvested, analysed and verifiable data
 - Improve **user experience** through visual presentation of data
 - Obtain more and more “**full digital**” products exploitable on a variety of platforms
- ◆ **Use a combination of the main features from the best practise**

- [illegible]

-

-

Architectural Overview

«Data Team»



«Customer»

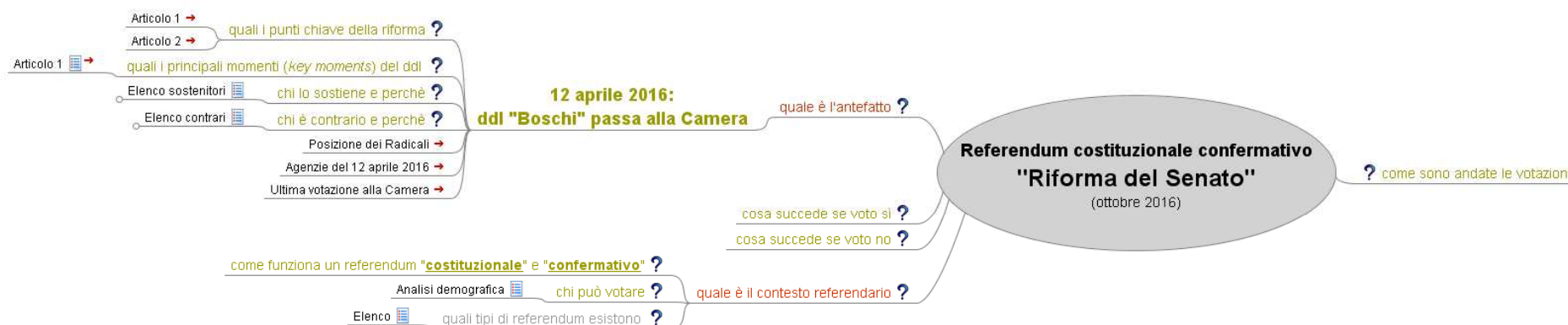
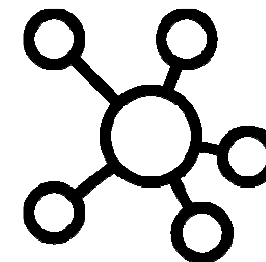




The Data Team

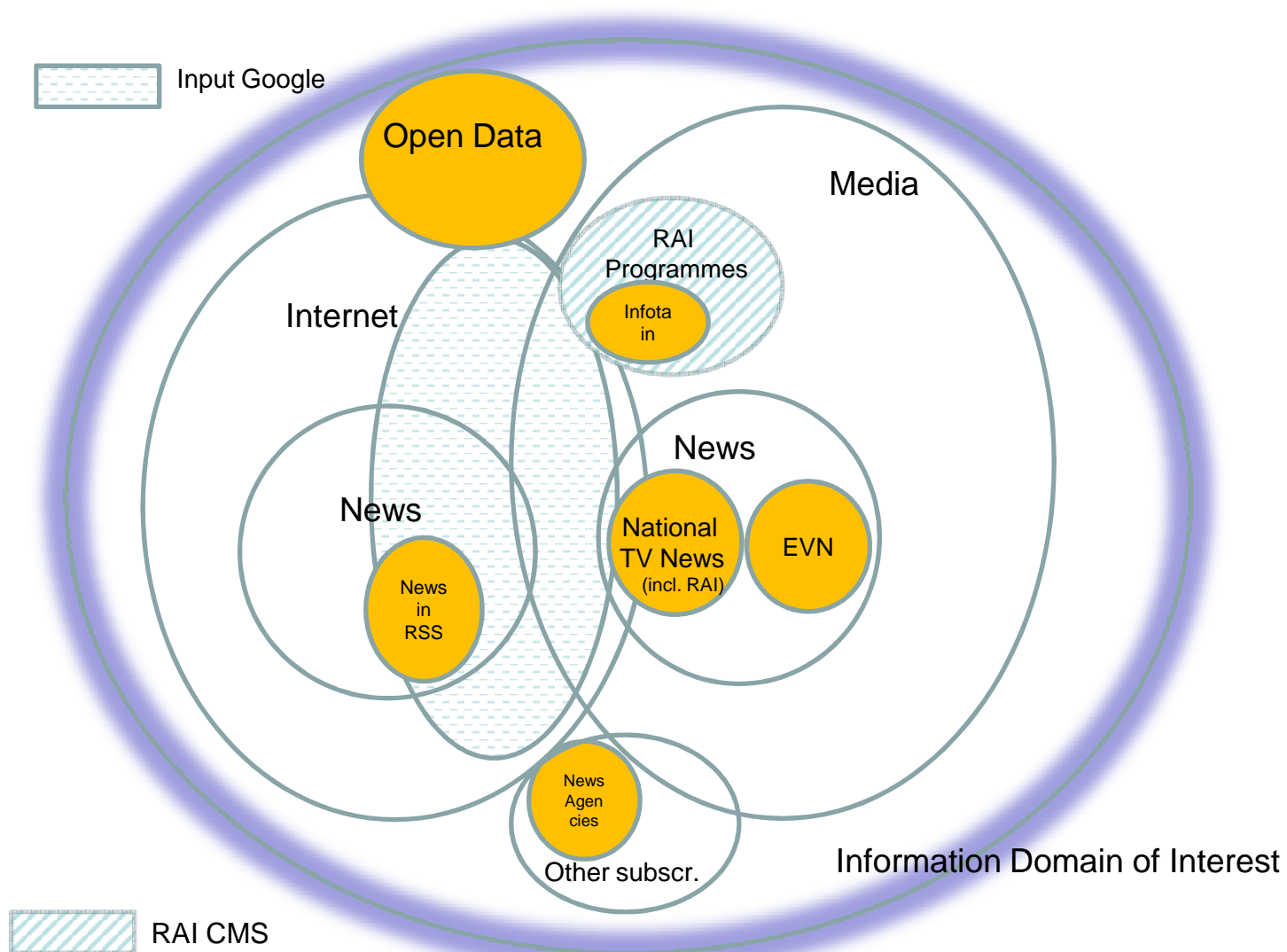


- ◆ **A group of people with multidisciplinary skills**
 - Journalistic skills
 - Technology skills
- ◆ **Agile interaction towards the objectives**
 - Technical members support journalists in finding / harvesting /processing data
 - Journalists set the editorial line, develop the story, assess data relevance
- ◆ **Result is that**
 - Efficiency is increased due to synergy and delegation
 - Cross-semination gives birth to unexpected explorations



<https://webbrain.com/brainpage/brain/434E72FE-3EED-7B13-2F44-561D8F294F28>

Which Information Sources?

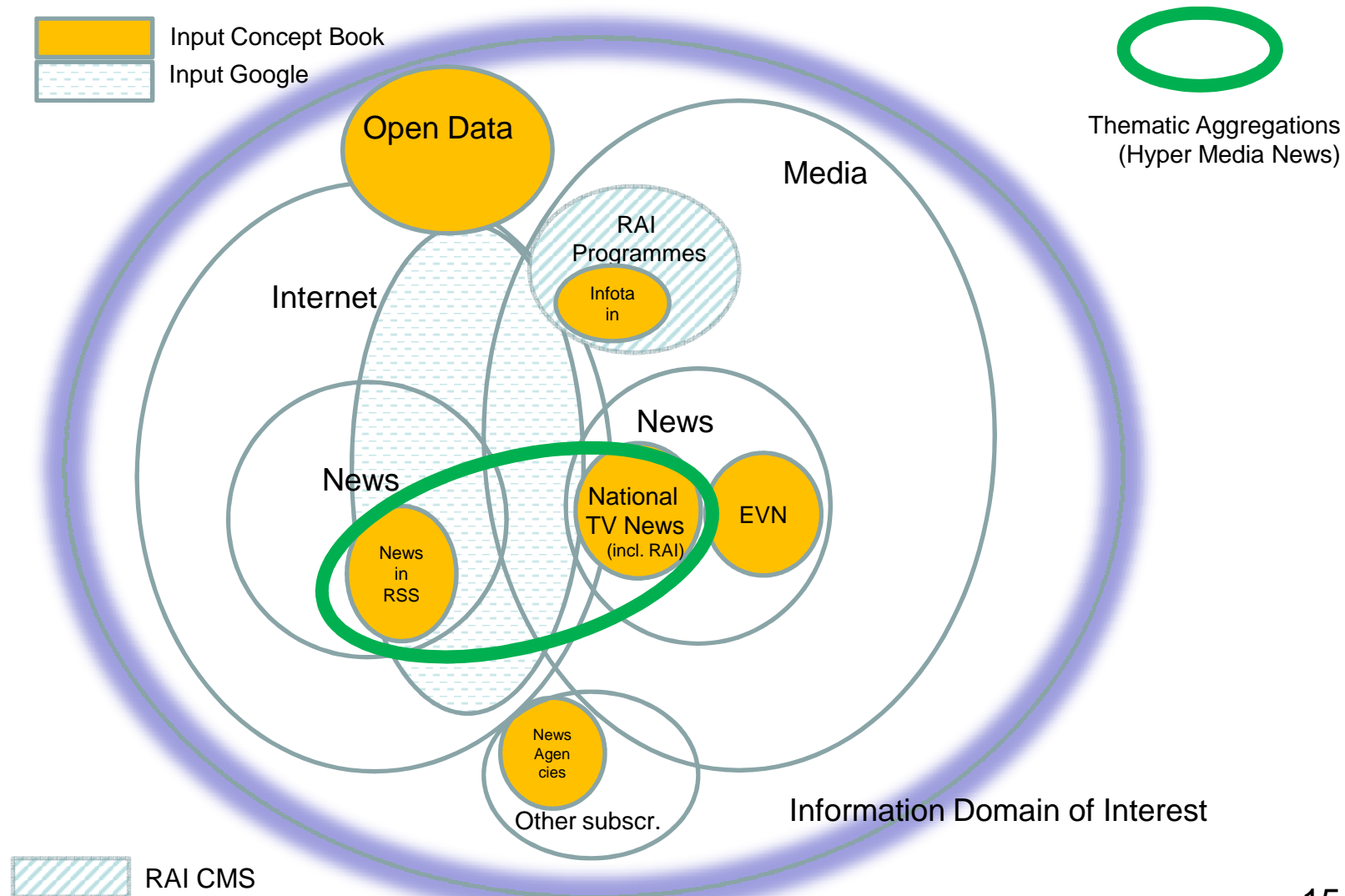


So what?

- ◆ **No one alone covers everything is needed for data journalism**
 - Neither Google
- ◆ **We need a wider approach than simply “search on the web”**
- ◆ **Solution:**
 - Integrated and flexible search and analysis of heterogeneous sources at enterprise level

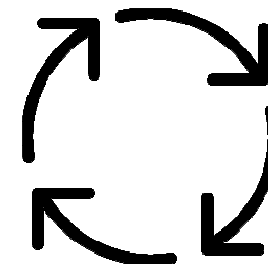
= RAI Concept Book

Which Information Sources?



- ◆ **Rai Concept Book is a portal for professional information services that addresses DDJ tasks with a uniform and holistic approach**
- ◆ **Artificial intelligence and advanced statistical tools are used to automate tasks such as information extraction and multimedia content analysis**
- ◆ **The system allows to define customized search profiles that are automatically and dynamically updated with the relevant contents found in the monitored information sources**

Processing Pipeline



1. Ingest

- RSS feeds
- Blogs
- DTT streams
- EBU Eurovision News
- Rai Archives
- News Agencies
- Open Data

2. Process

- Speech to Text
- Natural Language Processing
- Document Classification
- Named Entity Recognition

3. Understand

- News aggregation
- Topic identification
- Data warehousing

4. Archive & Access

- Indexing
- Search & Retrieval
- Browsing & filtering
- Recommendation
- Exportation



Personalised Dashboard



Registra un nuovo profilo di ricerca

Titolo del profilo
Es.: Expo Milano 2015

Termini di ricerca
Es.: expo,expo milano,expo 2015
Separare i termini di ricerca con una virgola

Invia

Seleziona un profilo tra quelli disponibili

Berlusconi Assolto	Crisi M.o.	Ebola	Educational	Expo Milano 2015	Isis
Istituzioni	Papa Francesco	Patto Del Nazareno	Piemonte	Sanremo 2015	
Spread	Terrorismo Internazionale	Turismo	Tutti I Contenuti		

Sviluppato da **Rai** Centro Ricerche e Innovazione Tecnologica



Sources

News
in
RSS

News
Agen
cies

Infotain

Open Data

Thematic Aggregations

EVN

National
TV News
(incl. Rai)

Profilo: 'Expo Milano 2015'

Filtri attivi

Termini di ricerca:

Expo Milano

Expo 2015

Expo Due Mila Quindici

Expo

☐ Includi tutti i termini della ricerca

Dal:

03/06/2015

A:

03/06/2016

☒ Ordina per rilevanza (dal più rilevante al meno rilevante)

☐ Ordina per data (dal più recente al meno recente)

Agenzie Stampa (archivio) Agenzie Stampa (live) Archivio Rai Catalogo Open Data Dossier Rai HMN (Hyper Media News) Notizie EVN (Eurovision News Exchange) Notizie TV

Notizie Web

Risultati 1 - 5 di 8566

Risultati per pagina

Tags

Expo

Primate Camera York Pangi

Milan Majolino Maurizio Carlo

Andrea Luca Mattarella Bazzani Fir

Pansa Paolo Inter Tag Expo Giuseppe Marino Ma

Tronca

Unione Per Renzi Italia Sala

Scelta

Italy Giubileo Usa

Albero Matteo Lomi

Putin Sergio Rai

Francia Stati Agi

Legia Rho Mario De u

Venezia Martina Roma Expo Pisapia

Parole Day Palazzo Margni Regione Ca

Obama Giuliano Francesco Berlusconi

Russia Salvini Giovanni Beppe Carda

Cala il sipario sull'Expo di Milano

★★★★★ - Il Secolo XIX - 31/10/2015, 19:59:40 - Cronaca

TAGS: Martina - Cerimonia - Astana -

... del dialogo e della ricerca. Abbiamo guardato al cibo come lingua comune dei popoli, che non vuol dire omologazione, ma rispetto delle culture e delle loro radici, scoperta di valori profondi. Sala: «È stata l'Expo della gente! Applauditissimo il commissario di Expo 2015 Giuseppe Sala. Che ha iniziato parlando di numeri: «Un'ora fa - ha detto - l'ultimo visitatore ha varcato l'ingresso...»(Continua)

Expo: cala il sipario su edizione storica

★★★★★ - ANSA.IT - 31/10/2015, 17:42:56 - Cronaca

TAGS:

...Con circa mezz'ora di ritardo rispetto all'orario annunciato, alle 17:35 sono stati chiusi al pubblico gli ingressi di Expo Milano 2015. La giornalista e conduttrice Rai Cristina Parodi ha dato il via alla cerimonia di chiusura ufficiale di Expo Milano 2015, presente il presidente della Repubblica Sergio Mattarella. La conduttrice sul palco ha ricordato i numeri del successo di Expo...»(Continua)

Expo, il Sudan festeggia il National day

★★★★★ - Il Corriere Della Sera - 27/7/2015, 15:20:57 - Esteri

TAGS: Sudan -

...Culla del Nilo e terra delle potenzialità agricole e alimentari. Questo il tema scelto dal Sudan per la sua partecipazione a Expo Milano 2015 e ribadito nel corso del National Day dedicato al Paese africano. Alla giornata ha preso parte il ministro del Commercio, Elsadig Mohamed Ali, che ha sottolineato l'importanza del settore agricolo nell'economia del Sudan, che rappresenta oltre il 35...»(Continua)

Profilo: 'Expo Milano 2015' Home Impostazioni

Filtri attivi

Termini di ricerca:

Expo Milano

Expo 2015

Expo Due Mila Quindici

Expo

☐ Includi tutti i termini della ricerca

Dal:

01/05/2014

Al:

25/05/2015

☒ Ordina per rilevanza (dal più rilevante al meno rilevante)

☐ Ordina per data (dal più recente)

Risultati 1 - 5 di 890

Risultati per pagina Esporta

1 2 3 > >>

TG1 ore 13:30 del 01 maggio 2015
 ★★★★★ - Rai1 - 1/5/2015, 13:32:20 - Cronaca - 00:03:45:10
 ...ricorda un minuto di silenzio d' dichiaro' ufficialmente aperta l' esposizione universale **expo milano due mila quindici** con il mondo della politica le istituzioni la societa' civile fuori timore delle contestazioni sono appesi i black bloc tutta europa il peggio ha seguito in diretta la cerimonia di inaugurazione di **expo due mila quindici** momenti emozionanti con centinaia di ospiti migliaia...(Visualizza il filmato)

TG2 ore 13:00 del 10 giugno 2014
 ★★★★★ - Rai2 - 10/6/2014, 13:14:41 - Economia Credito Finanza - 00:01:52:2
 ... delle quindici se parliamo da shanghai neanche per darsi che l' **expo due mila quindici** deve essere un' occasione per l' italia per raccontare se stessi a raccogliere renzi c' e' la comunita' italiana d' affari di shanghai sono circa mille i nostri imprenditori riuscita ad affermarsi nel mercato interno stesso tempo competitivo del imu venti ricorda che cinque anni fa in questo e' stato il padiglione italiano...(Visualizza il filmato)

TG1 ore 20:00 del 01 maggio 2015
 ★★★★★ - Rai1 - 1/5/2015, 19:57:39 - Individuo Famiglia Associazioni Società - 00:02:45:10
 ... cdu l' alimentazione nutrire il piano grande giornale latone finalmente giorno arriva sotto la pioggia ha accompagnato un messaggio di papa francisco il pontefice si fa portavoce dei poveri chiede pane di vita di lavoro per tutti portare la bandiera italiana una delegazione di dieci mila lavoratori che hanno preparato **expo due mila quindici** a chiudere la fila i rappresentanti del nepal paesi...(Visualizza il filmato)

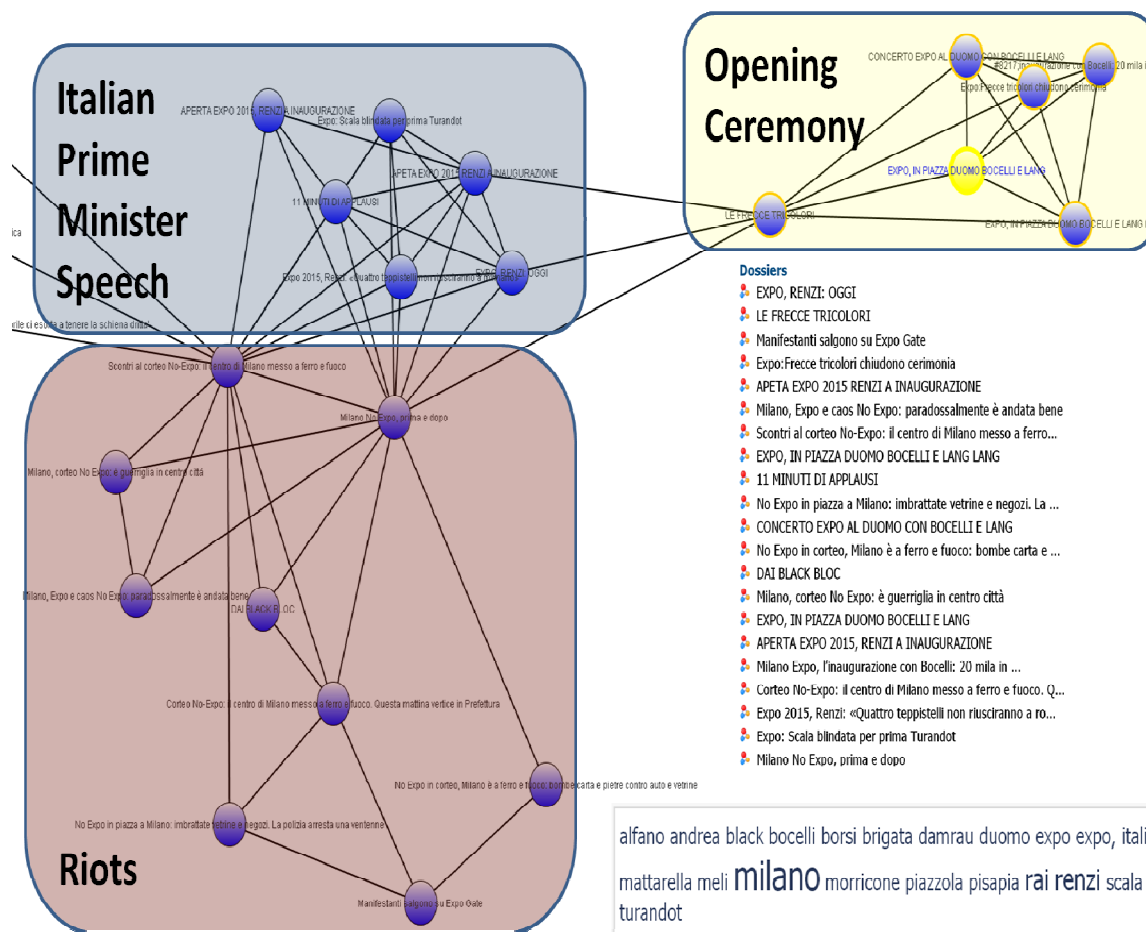
TG1 ore 13:30 del 10 giugno 2014
 ★★★★★☆ - Rai1 - 10/6/2014, 13:38:31 - Individuo Famiglia Associazioni Società - 00:01:47:5
 ...risultati scontati in un' intervista lewinsky comunque catturare il voto dei moderati che determinano seconda tappa della missione in oriente ferrara anzi dopo il vietnam nel paniere a shanghai occasione per parlare dell' **expo due mila quindici** come vetrina per chi vuole investire in italia e corrispondente claudio pagliara parenti un forte sostegno gli

Interactive charts of semantic entities



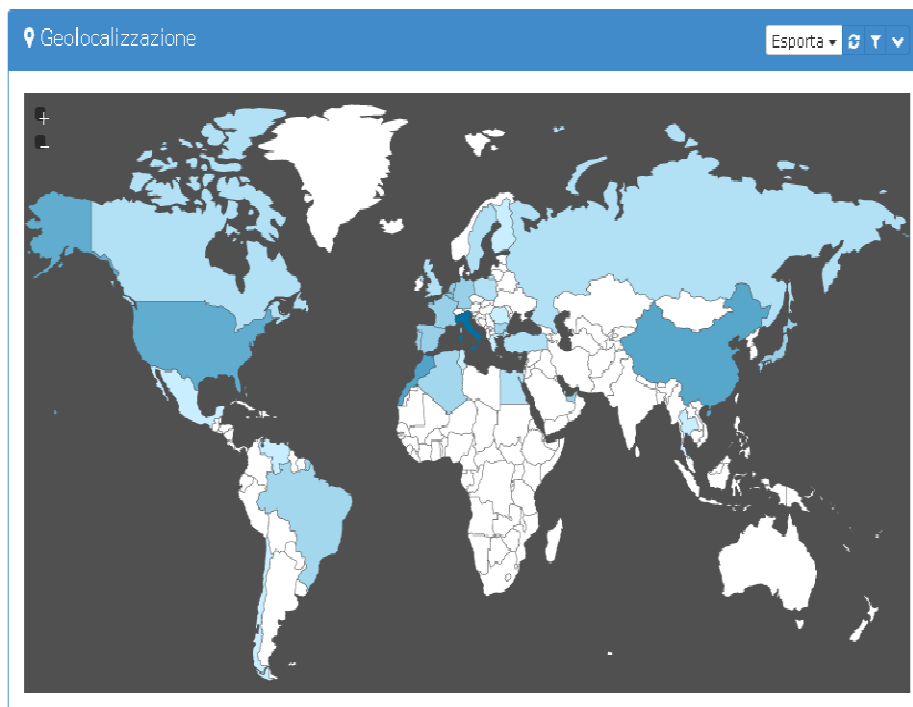
Who?

Exploring facts



What?

Localising in space

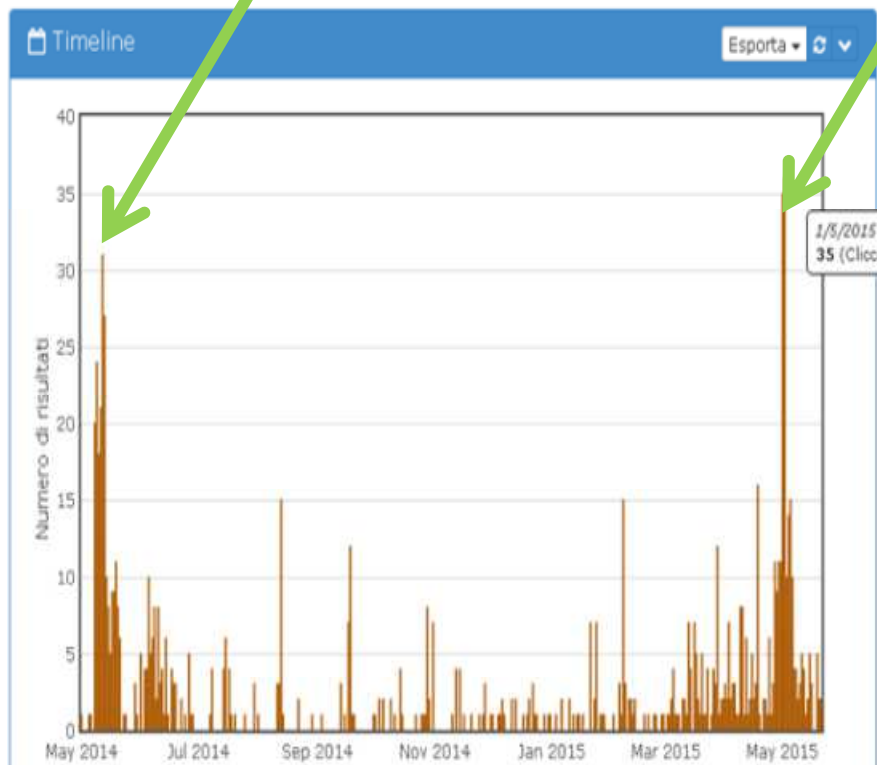


Where?

... and time

Scandal on public tenders for Expo works

Opening Ceremony



When?

A platform for the Data Value Chain

- ◆ **Comprehensive Knowledge Archive Network (CKAN)**
- ◆ **Open Source used by many organisations as a platform for open data publication**
- ◆ **Used in the project as a platform for data journalism production**
 - Editorial staffs \leftrightarrow CKAN organisations
- ◆ **Easy integration with other CKAN-based open data repositories**
 - Provides updates of data automatically
- ◆ **Extensible via plugins**



The screenshot displays the Rai .it website's data portal. The top navigation bar includes the Rai .it logo, links for ConceptBook, Dataset, Organizzazioni, Gruppi, and Informazioni, and a search bar. A user profile for Alberto Messina is visible in the top right corner. The main content area features a large graphic on the left with the text "Benvenuti* su dati.rai.it" and "Il portale di raccolta dati della RAI". To the right, there is a "Cerca i dati" section with a search input field containing "Per es. ambiente" and a search button. Below this, there are "Tag popolari" (popular tags) including "servizi", "dipendenti", and "citta-metropolitana...". Further down, a section titled "Regione Basilicata" contains two data-related items: "Qualità dell'aria. Dati di monitoraggio" and "Elenco beneficiari fondi PO FES dal 2008 al 30/4/2016".

- ◆ **Many many options available**
- ◆ **Default choice for Datawrapper in this phase of the project**
 - Open Source
 - Many graphics & many options
- ◆ **Extensible via plugins**

Rai.it Datamapper

Login / Sign Up

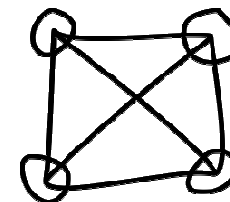
RAI - Data visualization

Il sistema di data visualizzazione





Integration



- ◆ **Source data selected from the RAI Concept Book become datasets in CKAN**
- ◆ **Datasets harvested by CKAN are classified and analysed by RAI Concept Book toolbox**
- ◆ **Integration between CKAN datasets and Datawrapper**
- ◆ **SSO between CKAN and Concept Book**



Some initial products



- ◆ **Finding the “right” approach at Data Journalism taking into account RAI’s peculiarities**
- ◆ **Multidisciplinarity is key**
- ◆ **Integration of proprietary and SotA tools works**
- ◆ **Still much work to do at all levels**
 - Integration
 - Workflow
 - Processes and skills
- ◆ **Very good feedback from RAI top level mgmt**

- ◆ **Tighter integration with Semantic Data**
 - Open data in RDF
- ◆ **Developing new automatic classification technologies**
 - On more relevant taxonomies than those currently used
- ◆ **Improve model for “concepts”**
 - Evolving towards more semantic structuring
- ◆ **Does Visual Search play a role in Data Journalism?**
- ◆ **Second screen & data journalism**
 - You’ll have a look at this in a few minutes



Thank You!

Contact me at

Dr. Alberto MESSINA, PhD
RAI – Radiotelevisione Italiana
Centre for Research and Technological Innovation
alberto.messina@rai.it