# Action MADONA, premiers résultats

Patrick Marcel, LIFAT, Université de Tours
Symposium Madics, 07/07/2021

# MADONA

*Maîtriser l'Analyse interactive de DOnnées pour la NArration journalistique*

**OÙ ?**

Datajournalisme,
Data sciences,
Storytelling,
Communication
Journalistes,
Data scientistes
Associations citoyennes

**POUR QUI ?**

Journalistes
Services de communication
Start-ups de service Open data
Chargés de com'
Etudiants en infocom
Associations citoyennes

**QUOI ?**

Articles de la PQR et données sources
Données locales *Open data*
Données hétérogènes (structurées ou non)
Données *How-to* des journalistes

**COMMENT ?**

Enquêtes quali. sur les pratiques et usages
Réunions pluri-disciplinaires de validation des modèles
Hackathons datascientistes /datajournalistes

**POUR QUOI ?**

Livre blanc des pratiques
Prototype d'aide à l'analyse et la rédaction
Meilleure compréhension des processus de construction de narration

**AVEC QUI ?**

Laboratoire d'informatique fondamentale et Appliquée de Tours
Centre d'études sur les médias, les technologies et l'internationalisation
Centre de recherche sur les médiations
Rue89 Strasbourg

**2019-2022**

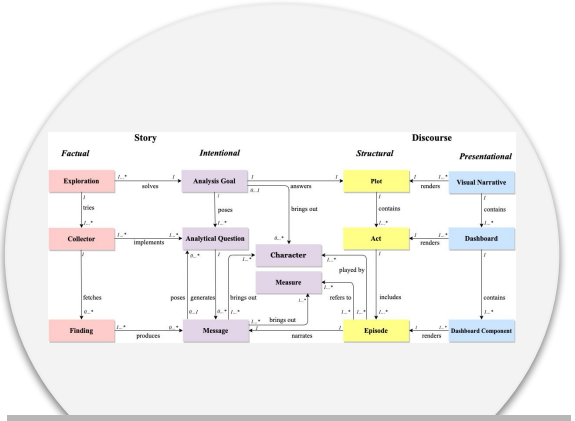
GARANTI SANS FAKE NEWS*!

https://www.madics.fr/actions/madona/

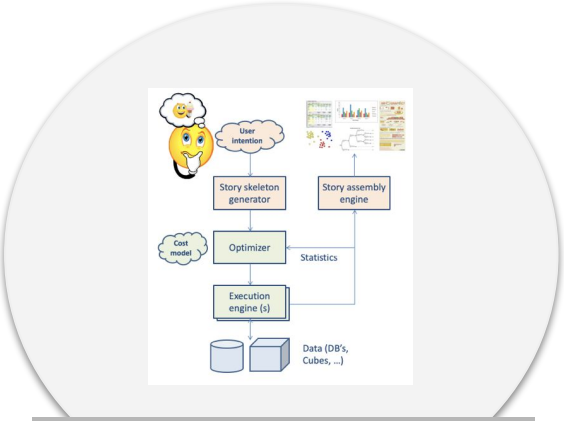

POUR QUOI ?
Meilleure compréhension des processus de construction de narration

# Accompagner la mise en narration des données

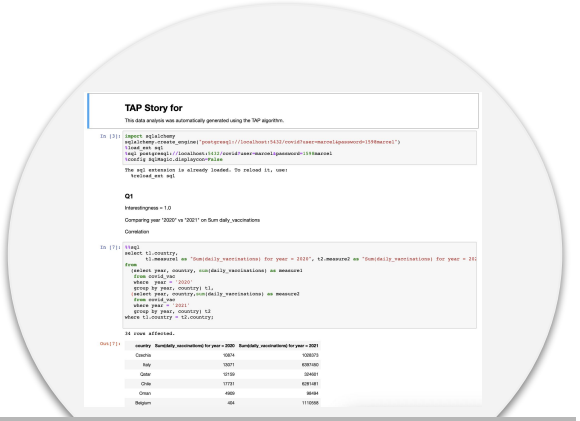
- Objectif : Identifier les méthodologies de construction de narration à partir de données
  - Analyses des pratiques socio-professionnelles et analyses des productions
  - Automatisation de l'exploration de données et aide à la construction de narrations




**Conceptualisation**

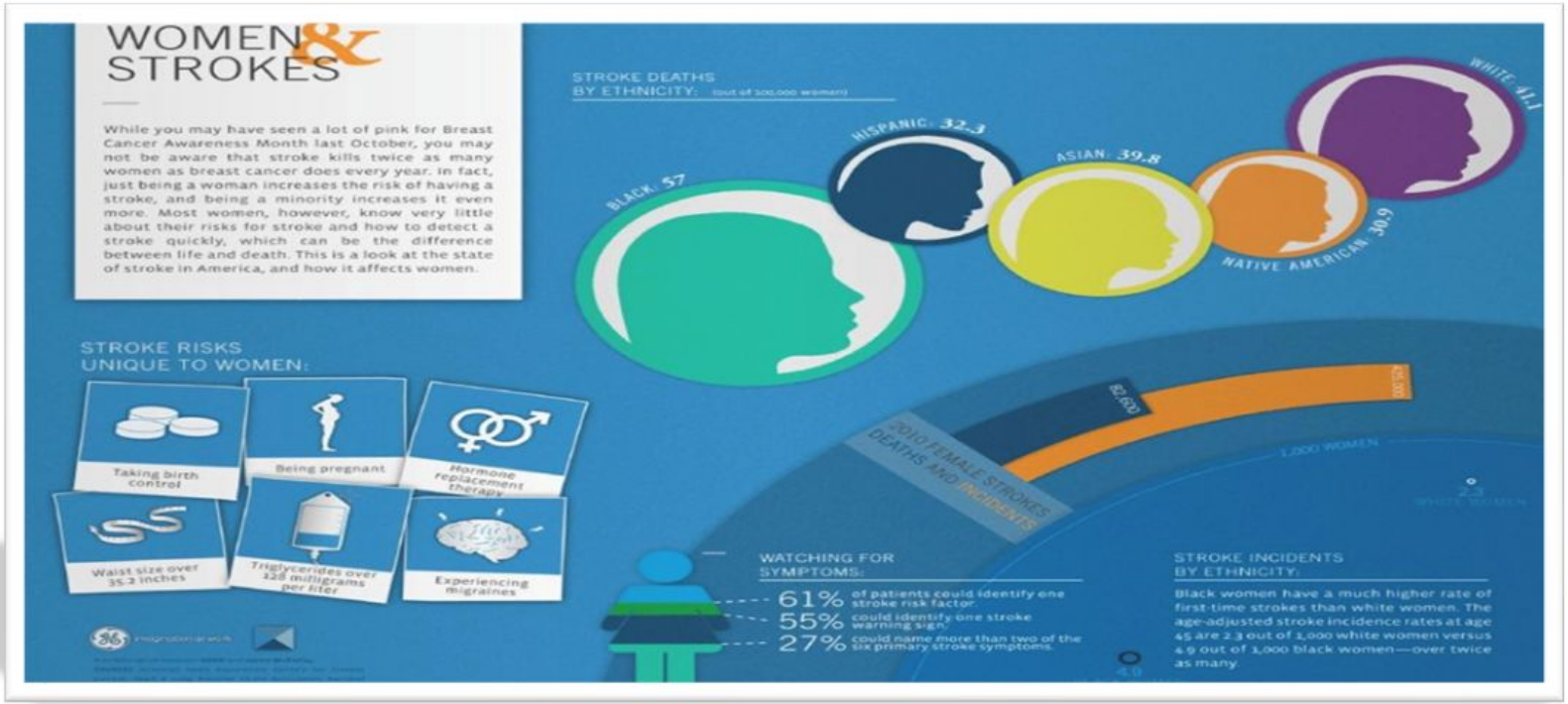



**Construire un vocabulaire**




**Génération de notebooks**

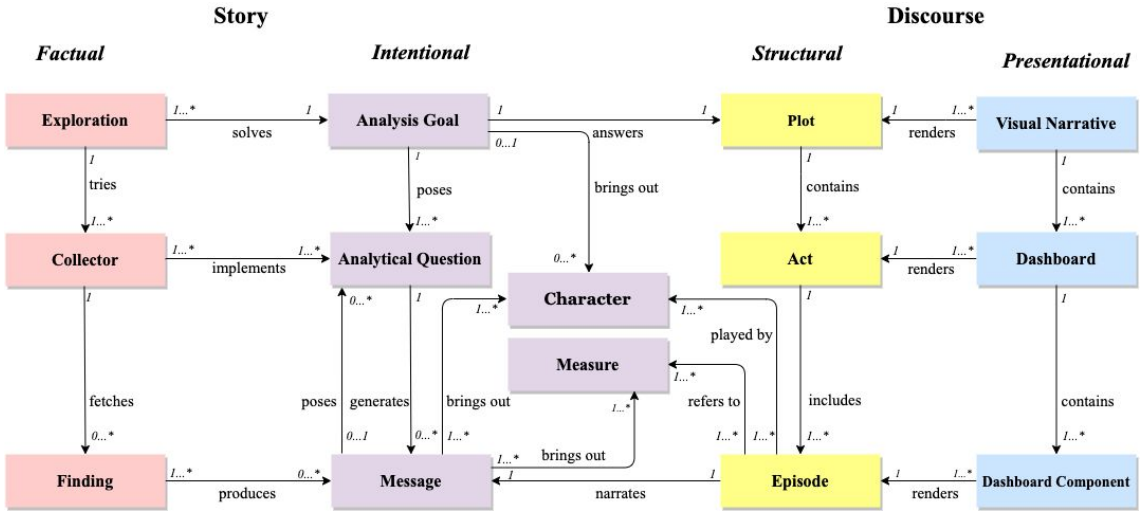
# Exemple de narration de données




*https://www.good.is/infographics/facts-about-women-and-strokes*

# Conceptualisation

Recensement des concepts et processus guidant la narration de données

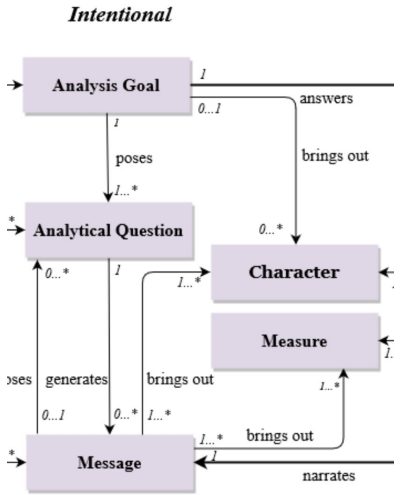


# Vocabulaire d’intentions

## Simplifier l’accès et la manipulation de données

## 5 primitives

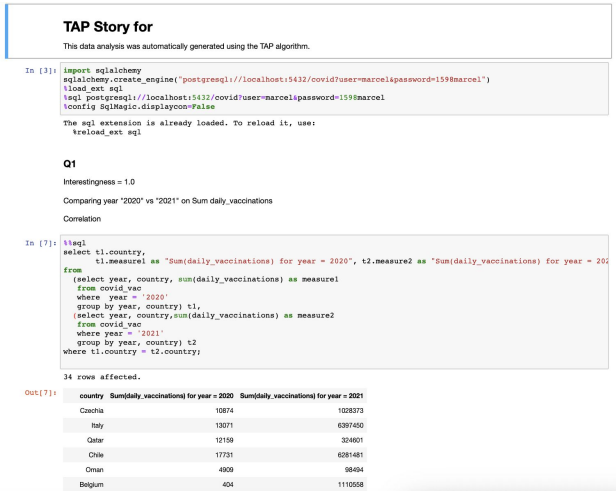
- **Description** (“**que** nous disent ces données ?”)
- **Comparaison** (“**comment** sont ces données par rapport à d’autres données ?”)
- **Explication** (“**pourquoi** ces données sont comme elles sont ?”)
- **Prédiction** (“comment seraient ces données **si...**?”)
- **Suggestion** (“que devrais-je savoir **d’autre** ?”)

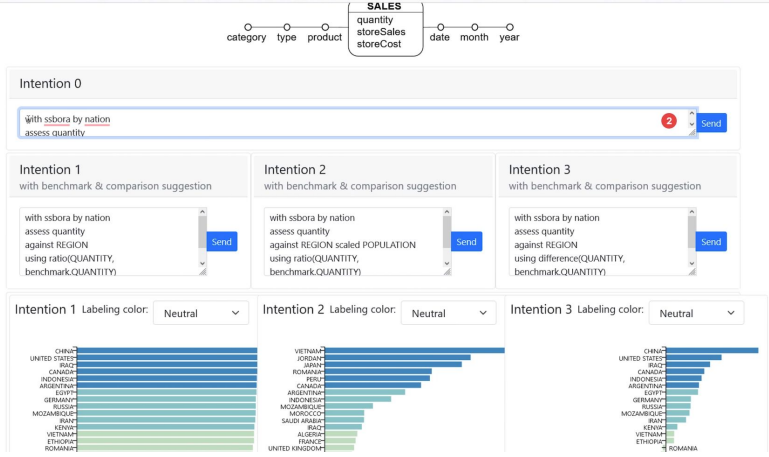


# Comparaison

2 prototypes de génération de notebooks de comparaisons :

1. Suite de comparatifs 1 contre 1, mise en évidence de faits statistiquement significatifs, prescriptif
2. Recommandation de comparaisons, mode automatique ou semi-automatique, nécessite un prétraitement en amont



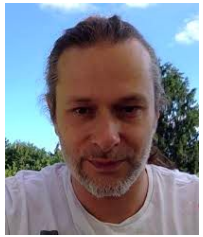



# Publications


1. Chagnoux M. (2020) "La datavisualisation, double point d'entrée du datajournalisme dans la PQR", Entre data journalisme et pratique infographique, Interfaces Numériques, Volume IX, n° 3/2020.
2. Faten El Outa, Matteo Francia, Patrick Marcel, Veronika Peralta and Panos Vassiliadis: A conceptual model of data narrative for exploratory data analysis. ER 2020
3. Faten El Outa, Patrick Marcel, Veronika Peralta, Matteo Francia and Panos Vassiliadis: Supporting the generation of data narratives. ER 2020 Demo. https://github.com/OLAP3/pocdatastorytelling
4. Alexandre Chanson, Ben Crulis, Nicolas Labroche, Patrick Marcel, Verónika Peralta, Stefano Rizzi, Panos Vassiliadis: The Traveling Analyst Problem: Definition and Preliminary Study. DOLAP 2020
5. Antoine Chédin, Matteo Francia, Patrick Marcel, Verónika Peralta, Stefano Rizzi: The Tell-Tale Cube. ADBIS 2020
6. Verónika Peralta: From source data to data narratives: accompanying users in the way to interactive data analysis. Keynote, ADBIS/TPDL/EDA Joint Conferences, virtual conference, August 2020. https://youtu.be/A_tOQcyKc-w
7. Matteo Francia, Matteo Golfarelli, Patrick Marcel, Stefano Rizzi, and Panos Vassiliadis. Assess queries for interactive analysis of data cubes. EDBT 2021

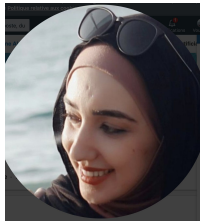
# Madona

Marie



Patrick

Raphael



Faten

Alex

Lucile

Nicolas

Veronika

Julien

Matteo
(Bologna)

Panos
(Ioannina)

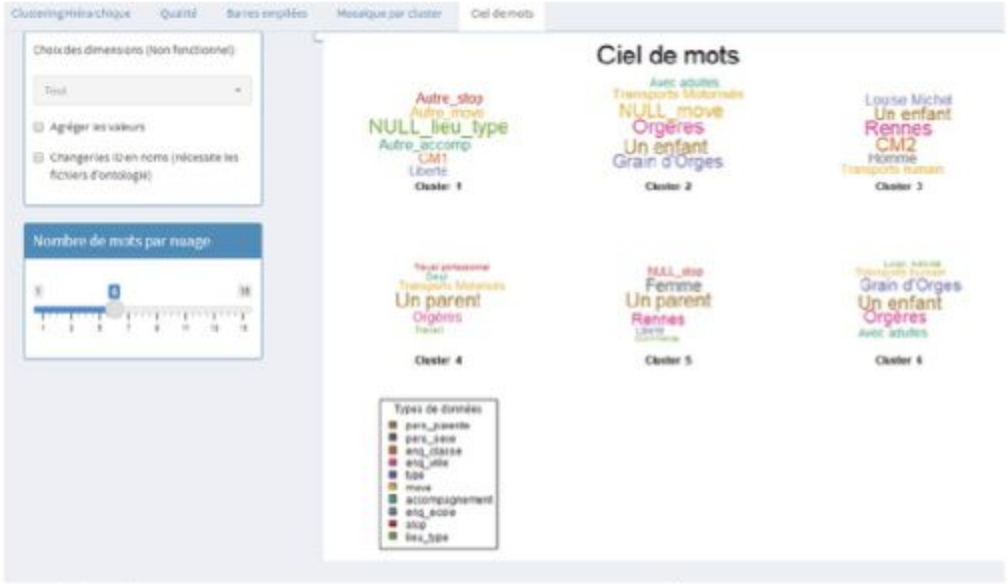
Stefano
(Bologna)

# ANR Mobi'Kids (CNRS, univ. Rennes, Grenoble, Tours)

- **Objectifs**
  - Étude de l'autonomie des enfants et des cultures éducatives
  - Développement d'un protocole méthodologique mixte
  - Zone d'étude : **Rennes et Orgères**
- **Données**
  - Recueil de données des activités et des mobilités + données socio-psychologiques (entretien)
    - Tracker GPS + enquêtes sur tablette (activités, mode de déplacement, type d'accompagnement des enfants…)
    - 4 271 séquences d'activités journalières d'enfants de CM1, CM2 et 6$^{eme}$ ou de l'un de leurs parents, recueil de 5 à 15 jours par an tous les 6 mois à 1-2 ans pour chaque individu
    - Données de parcours commentés
    - Données sensibles (CNIL)
  - Reprise des EMD (enquête ménage-déplacement) de Rennes
- **Processus de fouilles de données**
  - Analyses des séquences d'activités journalières
  - Définition de clusters de séquences similaires
  - Explicabilité des clusters (éléments saillants de chaque clusters)
  - …

ANR – MOBI'KIDS, ANR-16-CE22-0009
(Coord. Sandrine Depeau)

# ANR Mobi'Kids




- **Développement d'un outil d'analyse et de visualisation : SIMBA**
- **Exemple de connaissances extraites**
  - données Mobi'kids : Les comportements des enfants à Rennes et Orgères sont différents (mode de transport, accompagnement) :
  - données EMD : Il y aurait un gain en autonomie lors du passage en 6ième plus sensible pour les garçons que pour les filles (EMD Rennes)
- **Contacts**
  - thomas.devogele@univ-tours.fr
  - sandrine.depeau@univ-rennes2.fr

# EGC 2022
# BLOIS
# 24-28 janvier
# egc2022.univ-tours.fr

Sihem Amer-Yahia
Présidente du comité de programme

## Dates de soumission

Résumé : **8 octobre**
Article : **15 octobre**


https://twitter.com/egc2022