

# Média sociaux et Infrastructures de calculs & données

Jean-marc.francony@univ-grenoble-alpes.fr



JCAD2019 Toulouse



# L'engouement pour les médias sociaux

- L'ampleur des *usages* et des *pratiques sociales* associés aux médias sociaux suscitent un grand intérêt pour les communautés SHS;
- Cet intérêt tient:
  - en premier lieu, aux productions informationnelles et documentaires publiées;
  - en second lieu, aux potentialités des mécanismes de *traçabilité* appliqués:
    - aux actions du réseautage et aux interactions publiques entre les usagers,
    - aux comportements individuels et collectifs des usagers dans la mise en œuvre des services;
- Cet intérêt trouve sa justification dans l'étude:
  - des modalités de production,
  - des dispositifs info-communicationnels multicanaux,
  - des formes d'actions collectives et communautaires.

# Fabriquer ses données

- Dans le contexte RSN, l'approche empirique, guidée par les données, s'impose du fait d'une conception événementielle de la production des données.
- Dans cette approche, la constitution de collections de données brutes du Web (*raw data*) est en elle-même un cycle de fabrication, initialisant le cycle de vie des données.
- La fabrication des collections requiert des compétences une ingénierie de la donnée à *façon* nécessitant un accompagnement technique de proximité.

# Illusions et désillusions...

- L'accessibilité publique des média sociaux sur le Web entretient l'illusion d'une *cueillette* simplifiée des données;
- Le *scraping* des pages Web parait résoudre tous les problèmes de la constitution de jeux de données structurées (en l'absence d'API);
- Les conséquences techniques et méthodologiques du changement d'échelle de l'approche par les données sont peu anticipées ou mal appréciées.

# Achoppements...

- La transition entre le poste de travail personnel et l'environnement numérique en ligne se heurte aux difficultés de la compréhension technique:
  - de l'organisation virtualisée et distribuée du stockage de l'information,
  - des architectures logicielles et des niveaux de traitements.
- Les habitudes de travail avec les formats tabulés sont à l'origine de représentations linéaires des données, peu compatibles avec les réalités semi-structurées des données du Web.

# Constats...

- Les transformations techniques nécessaires au changement d'échelle constituent des verrous dans la mise en œuvre de projets SHS impliquant des données du Web.

# Pistes...

- Les choix techniques de virtualisation doivent être pensés globalement jusqu'au poste de travail afin de préserver la continuité des usages (gestion de fichiers, etc.);
- Les changements d'architecture souvent nécessaires pour amorcer le changement d'échelle doivent faire l'objet d'un accompagnement de proximité;
- Dans une phase transitoire (?), l'ingénierie des données doit être accompagnée au plus tôt, dès la définition de projet.