

Groupe de contact FNRS

« Analyse critique et amélioration de la qualité de l'information numérique »

18 mai 2022

« Apprentissage automatique : les challenges de la qualité des données dans la perspective d'une adéquation aux usages »

par

Laurence Dierickx

Introduction

Isabelle Boydens



MA^{en}
STIC

Master en sciences
et technologies de l'information
et de la communication

Groupe de contact FNRS

« Analyse critique et amélioration de la qualité de l'information numérique »

- Qualité des données :
 - « *fitness for use* » : adéquation des données à leurs objectifs (sociaux, industriels, juridiques, scientifiques, ...), sous *contrainte de budget*
 - pas de « *qualité totale* »
 - approche continue et pluridisciplinaire
- Groupe de contact FNRS créé en 1994 (25 ans en 2019)
 - <https://www.frs-fnrs.be/fr/financements/mobilite-fnrs/groupe-de-contact/91-nos-financements/mobilite-fnrs/groupe-de-contact/103-sciences-appliquees>
 - <https://www.frs-fnrs.be/fr/financements/mobilite-fnrs/groupe-de-contact/91-nos-financements/mobilite-fnrs/groupe-de-contact/104-sciences-humaines-et-politiques>
- Centre de compétence en qualité des données, Smals
 - <https://www.smals.be/fr/content/data-quality>
 - <https://www.smals.be/nl/content/data-quality>
- Cours de Master en STIC à l' Université libre de Bruxelles :
« Qualité de l'information et des documents numériques »
<https://www.ulb.be/fr/programme/2020-stic-b510>

Enjeux stratégiques lorsque l'information est un instrument d'action sur le réel

Quelques exemples :

- 1442, Laurent Valla démontre que la « *Donation de Constantin* » est un faux antidaté de 4 ou 5 siècles
- Étude du réchauffement climatique (couche d'ozone, 1980)
- Première guerre du Golfe (1990-1991)
- 1999, guerre du Kosovo, bombardement de l'ambassade de Chine à Belgrade
- 2008, 2017 interactions entre données et marché financier : Google Finance, Yahoo! Finance, World-Check (2017)
- 2013, secteur administratif (Obamacare)
- 2016, USA : coûts de la « non-qualité » (3.100 milliards \$, 20% PIB USA, T. Redman)
- 2017, domaine des pipelines, de l'hydrologie, ...
- 2018-2022, bases de données en matière de terrorisme, médicales (Implant files, Covid19, ...), Registre National, ...

Thématiques 2014 – 2020 (1)

- 2014

- « *Du stemma codicum au data tracking* » : vingt ans de recherche en matière d'évaluation et d'amélioration de la qualité des bases de données », par I. Boydens,
- « *Email Address Reliability*, » par V. Berten
Programme, slides et vidéo : <http://mastic.ulb.ac.be/2013/10/reunion-du-groupe-de-contact-fnrs-analyse-critique-et-amelioration-de-la-qualite-de-linformation-numerique/>

- 2015 : « *Ouverture des données, gouvernance et engagement citoyen: questions de qualité* »

- « *Gouvernance et individus: réflexion sur un couple impossible* », Par F. de Smet
- « *Semantic Information Technology as a provider of tools for a productive and efficient organization of government* » par Agis Papantoniou
Programme et slides : <http://mastic.ulb.ac.be/2014/10/reunion-du-groupe-de-contact-fnrs-analyse-critique-et-amelioration-de-la-qualite-de-linformation-numerique-2/>

- 2016

- « *Analyzing the evolution history of data-intensive systems in support to software maintenance*», par A. Clève
Programme, slides et vidéo : <http://mastic.ulb.ac.be/2016/02/reunion-du-groupe-de-contact-fnrs-analyse-critique-et-amelioration-de-la-qualite-de-linformation-numerique-3/>

- 2017

- « *Temps réel, déterminisme et ordonnancement : trois défis majeurs des systèmes embarqués*», par J. Goossens
Programme, slides et vidéo : <http://mastic.ulb.ac.be/2017/02/reunion-du-groupe-de-contact-fnrs-analyse-critique-et-amelioration-de-la-qualite-de-linformation-numerique-4/>

Thématiques 2014 – 2021 (2)

• 2018

- « *La préservation du patrimoine scientifique à l'heure du numérique* », par A. Leroy

Programme, slides et vidéo : <http://mastic.ulb.ac.be/2017/09/reunion-du-groupe-de-contact-fnrs-analyse-critique-et-amelioration-de-la-qualite-de-linformation-numerique-5/>

• 2019

- « *Méthodes quantitatives, analyse critique de données et histoire médiévale : des données contemporaines aux sources anciennes* », par S. de Valeriola

Programme, slides : <http://mastic.ulb.ac.be/2018/08/reunion-du-groupe-de-contact-fnrs-analyse-critique-et-amelioration-de-la-qualite-de-linformation-numerique-6/>

• 2020

- « *I want to talk to a HUMAN!* » : *impact de la qualité des bases de connaissances sur les agents conversationnels* », par M. De Wilde

Programme, slides : <http://mastic.ulb.ac.be/2019/09/reunion-du-groupe-de-contact-fnrs-analyse-critique-et-amelioration-de-la-qualite-de-linformation-numerique-7/>

• 2021

- « *Avez-vous essayé l'ignorance ? Les défis du knowledge management aujourd'hui* », par F. Rossion

Programme, slides : <https://mastic.ulb.ac.be/2021/03/reunion-du-groupe-de-contact-fnrs-analyse-critique-et-amelioration-de-la-qualite-de-linformation-numerique-8/>

Thématique 2022

**« Apprentissage automatique :
les challenges de la qualité des données dans la
perspective d'une adéquation aux usages »**

par Laurence Dierickx

Laurence Dierickx

Laurence Dierickx est docteure en Sciences de l'Information et la Communication, auteure d'une thèse traitant de l'automatisation de la production d'informations, soutenue en 2020, et titulaire d'un master en Sciences et Technologies de l'Information et de la Communication.

Elle est actuellement consultante indépendante en information numérique, collaboratrice scientifique du centre de recherche (ReSIC) et membre du Laboratoire des pratiques et identités journalistiques (LaPIJ) de l'Université Libre de Bruxelles. Elle enseigne également le journalisme de données dans le cadre du Master en Journalisme de l'ULB.

<https://difusion.ulb.ac.be/vufind/Author/Home?author=Dierickx,%20Laurence>

Programme 2022

- 14h Introduction groupe de contact FNRS (I. Boydens)
- 14h05 « Apprentissage automatique : les challenges de la qualité des données dans la perspective d'une adéquation aux usages » (L. Dierickx)
- 15h15 Table ronde (modérateur : S. de Valeriola)

Les slides des présentations seront disponibles en ligne sur le site de l'invitation à la fin de la rencontre FNRS.