

Gestion de la qualité des données ouvertes liées

—

État des lieux et perspectives

Delphine Barrau*, Nathalie Barthélémy**, Zoubida Kedad***, Brigitte Laboisse****,
Sylvaine Nugier‡, Virginie Thion‡‡

* Conix Consulting, France

Delphine.Barrau@conix.fr

** Gaz Réseau Distribution France (GRDF), France

Nathalie.Barthelemy@grdf.fr

*** Université Versailles Saint Quentin, DAVID Lab, France

Zoubida.Kedad@uvsq.fr

**** Base for Data Quality Systems (BDQS), France

blaboisse@bdqs.com

‡ Électricité de France (EDF) R&D, France

Sylvaine.Nugier@edf.fr

‡‡ Université Rennes 1, IRISA, France

Virginie.Thion@irisa.fr

Résumé. Sous l'impulsion de l'apparition de nouvelles technologies permettant la publication et l'exploitation de données ainsi que de contraintes réglementaires imposant à certaines entreprises et institutions d'ouvrir leur données, la publication de données liées est devenue un phénomène en pleine croissance. Cette immense ressource de données offre de grandes possibilités d'exploitation. Cependant, on constate un niveau très disparate de qualité des données publiées, rendant leur utilisation difficile, voir risquée. L'évaluation et la maîtrise de la qualité de ces données deviennent de ce fait un enjeu de premier plan. Dans cet article, nous dressons un état de l'art des approches méthodologiques et techniques de gestion de la qualité des données ouvertes liées proposées dans la littérature. Le périmètre couvert inclut les dimensions et métriques, les frameworks de gestion, les plateformes et outils associés, et enfin les cas d'étude de publication et d'utilisation des données ouvertes liées centrés sur la qualité de celles-ci. En nous fondant sur cet état de l'art, nous exhibons des problèmes ouverts et des perspectives de recherche du domaine.