# Dredging the data lake

October 10 2018, by David Bradley

Credit: CC0 Public Domain

Data lakes allow information to be added to a system without pre-processing or modelling. Contrast this with a conventional database where data must be delivered in a much more refined and formal manner. Thus a data lake offers much timelier speed of entry. However,

as research from Brazil shows, even though a data lake preserves highest granularity level of the data, that useful flexibility can be problematic too. "If not managed, it is easy to lose control of the repository because of the volume it holds and its growth," the team explains.

The researchers explain further that data lakes carry none of the semantics of a conventional database, but while this can be advantageous in avoiding certain types of bias when re-extracting and analyzing days, it does mean that understanding the contents of the data lake can become a rather cumbersome task. This, the team suggests, has perhaps undermined the widespread adoption and use of data lakes within the corporate environment and stymied acceptance of this useful tool because of certain misconceptions regarding how they might be used in data science efforts.

The team has now turned to knowledge management models to help them address the issues associated with data lake use and to enrich the data floating within to enhance information usability. They also add that through the use of a data portal platform and associated metadata they reason that their approach would provide easy access to the data lake maintaining and boosting its usefulness and precluding its denigration into a so-called data swamp.

**More information:** Jano Moreira De Souza et al. Using knowledge management to create a Data Hub and leverage the usage of a Data Lake, *International Journal of Knowledge Management Studies* (2018). DOI: 10.1504/IJKMS.2018.10015483

Provided by Inderscience

Citation: Dredging the data lake (2018, October 10) retrieved 8 July 2024 from

https://phys.org/news/2018-10-dredging-lake.html

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.