

# Angela Bonifati

Full Professor, Lyon 1 University

## Quality-driven Analytics on Scientific Data

Wednesday, April 14, 2021  
11:30am – 12:45pm Pacific time



Scan or visit URL to register:  
<https://data.cs.sfu.ca/rWap>



### Abstract

One of the key principles of data analytics in data science pipelines is that the quality of the results is dependent on the quality of the input data analyzed. Data collection and acquisition in scientific domains, such as for instance in healthcare and life sciences domains, tend to often introduce errors into the data, such as violations of business rules, typos, missing values and replicated entries. Moreover, data collected for the patients' signals might exhibit peculiar features that need to be taken into account within the analytical and inference processes.

I will present our latest results on enhancing the quality of querying and inference processes on scientific data and beyond. Among the others, we operate on real-life data of patients from several hospitals in the EU and provide the domain experts with useful data management and learning techniques that can help them with their diagnoses and analyses. First, inconsistency-aware annotations can enhance the data input to analytical processes. These annotations are further exploited during query processing in order to enhance the output of queries with inconsistency degrees. Second, feature-based similarities among time series corresponding to patients' signals help to better identify groups of patients and to assess their risks. Third, logic-based declarative privacy-preserving data integration allows to migrate clinical data from one hospital to another while ensuring privacy guarantees. In all cases, our research aims at providing the caregivers with a better understanding of their clinical data thanks to the improved outcomes of the performed analytics. In the talk, I will overview the research performed in my group on this topic as well as the related work and the future directions of investigation.

### Biography

Angela Bonifati is a Professor of Computer Science at Lyon 1 University and affiliated with the CNRS Liris research lab since 2015. In 2019 and 2020, she has been on leave at INRIA. Prior to that, she was working as a Professor at Lille 1 University (2011-2015) and as a researcher at CNR, Italy since 2003. She received her Ph.D. from Politecnico di Milano in 2002 and right after she was a postdoctoral researcher at INRIA Rocquencourt in Paris for one year. Her current research interests are on the interplay of relational and graph-oriented data paradigms, particularly on data integration and Big Data curation for life sciences, query processing and learning for structured and unstructured data models. She is the Principal Investigator of the ANR research project QualiHealth: Enhancing the Quality of Healthcare Data (2018-2023) and was the recipient of the IDEX Pulse Impulsion Individual grant in 2016. She is involved in several grants at Lyon 1 University, including a EU H2020 grant on data quality and data integration in a clinical data lake. She has also co-authored several publications in first-rate venues of the data management field along with two books (edited by Springer in 2011 and Morgan&Claypool in 2018) and an invited paper in ACM Sigmod Record 2018. She is the Program Chair of ACM Sigmod 2022 and was the Program Chair of EDBT 2020. She is Associate Editor of PVLDB in 2020-2022 and Associate Editor of IEEE ICDE 2021, 2018 and 2011. She is Associate Editor of the VLDB Journal, ACM TODS, Distributed and Parallel Databases and Frontiers in Big Data. She is the President of the EDBT Executive Committee and a member of the ICDDT council. She holds many visiting scholar positions in foreign universities in both Europe and North America, the latest of which at the University of Waterloo (Canada) in 2019. She is currently Adjunct Professor in the latter university.