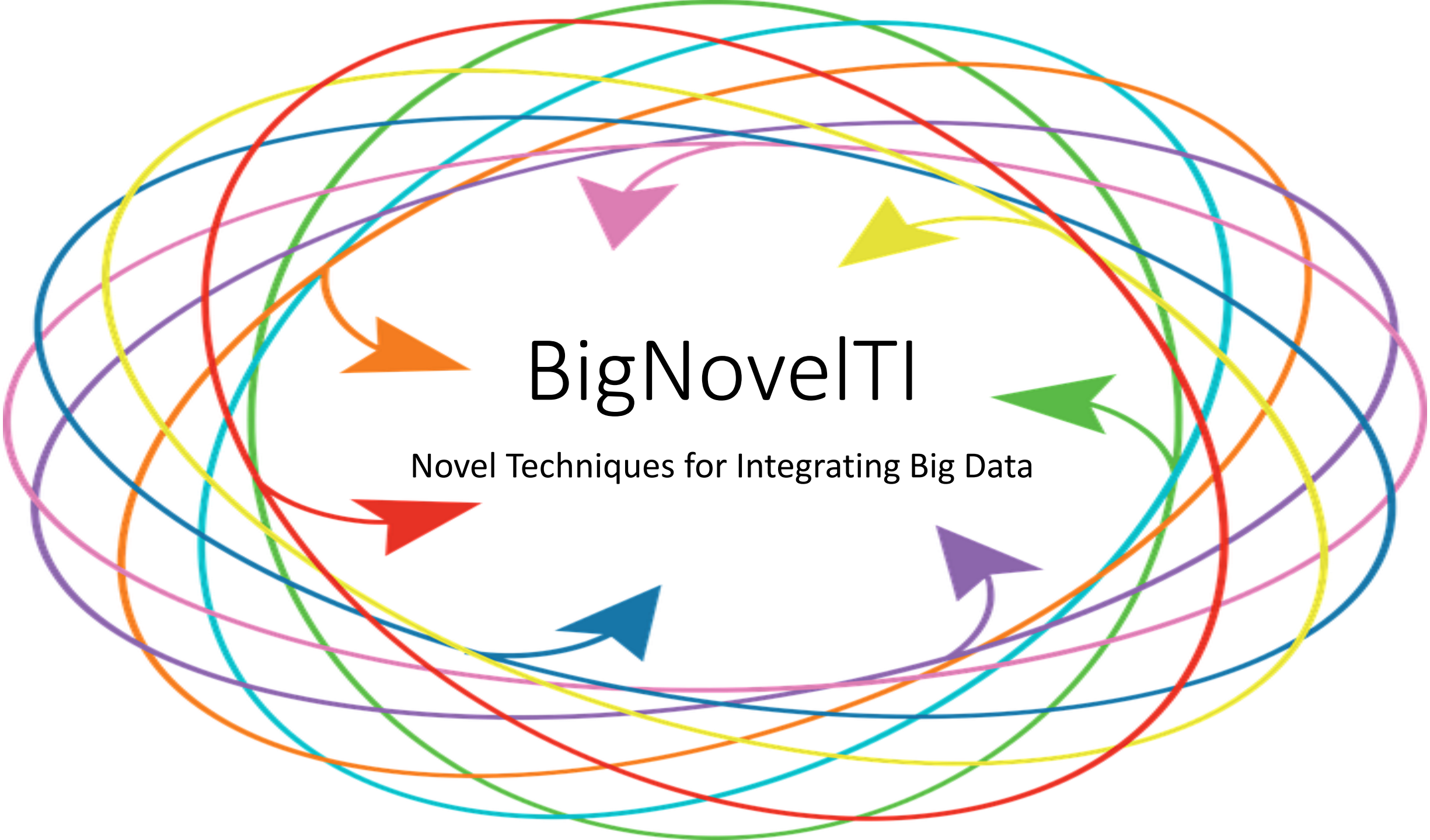

BigNovelTI

Novel Techniques for Integrating Big Data

Accepted Papers and Agenda

09:00-09:05: Welcome and motivation for the workshop

09:05-10:35: Session 1

- A Framework for Temporal Ontology-based Data Access: A Proposal
  - Sebastian Brandt, Elem Guzel Kalayci, Vladislav Ryzhikov, **Guohui Xiao** and Michael Zakharyaschev
- Detecting Semantic Correctness of Summarization in Self-Service Business Intelligence
  - Luis Ibanez-Gonzalez, **Jose-Norberto Mazon** and Elena Simperl
- Theta Architecture: Preserving the Quality of Analytics in Data-Driven Systems
  - **Vasileios Theodorou**, Ilias Gerostathopoulos, Sasan Amini, Riccardo Scandariato, Christian Prehofer and Miroslaw Staron

10:35-11:00: Coffee break

11:00-12:30: Session 2

- Spatio-Temporal Evolution of Scientific Knowledge
  - **Goce Trajcevski**, Xu Teng and Shailav Taneja
- Panel on Novel Big Data Integration Techniques - What is New?
  - Panelists: Robert Wrembel (Poznan University of Technology), Guohui Xiao (Free University of Bozen-Bolzano).
  - Moderator: Oscar Romero (Universitat Politècnica de Catalunya)

12:30-14:00: Lunch break

Panel on Novel Big Data Integration Techniques - What is New?

- Is classical data integration different from data integration for Big Data?
- If so, what are the hot topics in this area?

Panel on Novel Big Data Integration Techniques - What is New?

- Data Warehousing (DW) has been the in-company de-facto standard for data integration
- Big Data poses several challenges due to the autonomy of the sources and the uncertainty of the requirements for data integration. For this reason:
  - Bottom-up data integration approaches, performing a virtual integration, are a good fit for integrating disparate autonomous and heterogeneous data sources in large-scale distributed decision support systems, a current trend to contextualise the in-house data.
  - However, virtual data integration tends to suffer from poor performance, which hinders the right-time analysis approach, and needs to be adapted and combined with materialisation to meet the new requirements brought by Big Data.
  - Variety and and the need to deal with external non-controlled sources in an automatic way require to look at this problem from a broader perspective than the one of traditional data management, and semantics need to be included in the data integration processes. In such setting, domain knowledge is represented in an ontology, and inference over such knowledge is used to support data integration.
  - However, this approach poses significant computational challenges that still need to be overcome, and hence its potential has not been fully exploited yet in real world applications.

For these reasons, (i) providing different degrees of coupling of the integrated data sources based on their heterogeneity and autonomy, and (ii) dealing with and integrating semantics as a first-class citizen are open questions for novel scenarios.

Panelists

- Robert Wrembel (Poznan University of Technology) and
- Guohui Xiao (Free University of Bozen-Bolzano)
- Moderates: Oscar Romero (Universitat Politècnica de Catalunya)