Research Associate: Basic Grade  
Selection No.: CRN15CSS5

Sectoral Scientific Commission (CSS):  
CSS5 “science of data and models”

**Title**  
Methodological developments in data science, applied to massive, multimodal and heterogeneous data.

**Keywords / disciplines**  
Mining of heterogeneous data, data lakes, semantic web, ontology, knowledge representation and visualisation.

**Context and Issues of the Global South**  
The increasing size, scale and number of data sets from observation, experimentation and digital simulation require new methods to be defined, enabling interoperability, data integration and knowledge extraction within these mass of heterogeneous, multi-modal and silo-managed data. This in order to be able to 1) enable spatio-temporal and thematic representations of data, compatible with cross-disciplinary analyses, 2) ensure the reproducibility of analyses (modelling, simulation, visualisation) involving the combination of multidisciplinary data or algorithms, and lastly 3) provide visual and interactive data exploration to produce new interpretations. This has created a new and rapidly expanding field of research in itself. The IRD wishes to strengthen expertise in this field while positioning it around the objectives of Sustainable Development Goals and the science of sustainability wherever it is involved.

**Current practices and needs**  
The applicant will bring methodological expertise in the extraction, representation, and reconciliation of cross-disciplinary knowledge, and in the interactive, graphical visualisation of massive multimodal and heterogeneous data sources. He or she will contribute to design innovative approaches that will use and integrate methods and techniques from knowledge representation, data mining and data visualisation that can be applied to large structured and cross-disciplinary databases, the semantic web and data lakes. Researcher profiles showing the development of research on 1) knowledge formalisation and representation, and 2) data mining methods in large data sets will be particularly appreciated. Examples of IRD topics that relate to issues around interoperability, data integration and knowledge extraction include the environment, health, and social sciences for which IRD develops its research to drive development.

**Desired skills / abilities**  
Research experience in formal knowledge representation, data mining in heterogeneous and multimodal datasets, visualisation of complex data, and publications in leading conferences and journals in the field. Interest or experience in sustainable science projects is especially desired. This profile spans the main areas covered by the Sustainable Development Goals.

The IRD Hiring Policy Promotes Gender Equality in the Workplace