

# SEMINAR

**Speaker: Prof. Arek Kasprzyk**

Center for Translational Genomics and Bioinformatics,  
San Raffaele Scientific Institute,  
Milan, Italy

After earning a medical degree and a PhD in molecular biology, Arek Kasprzyk decided to pursue his passion for information technology and obtained an MSc in Bioinformatics. This unique background enabled him to obtain a position at the European Bioinformatics Institute in Hinxton, United Kingdom, where he designed and implemented BioMart, an innovative open source software for biomedical research, which provided the first large scale federated data management solution. He was personally recruited by the Ontario Institute for Cancer Research (OICR) in Toronto, Canada to lead the International Cancer Genome Consortium Data Coordination Centre, and create the architecture to manage their data, which will eventually be equivalent to 50,000 Human Genome Projects. Kasprzyk currently holds the position of Head of Data Management at the Center for Translational Genomics and Bioinformatics at San Raffaele Research Institute in Milan, Italy. He provides data management support for the institute's initiatives and oversees the development of tools for analyzing next generation sequencing data.

**Date:** Monday, April 25, 2016

**Time:** 1:00 PM

**Venue:** Engineering Building, Second floor,  
Room 24C28 (ECE Seminar Room)

## Title

### Data management for large collaborative projects: challenges and solutions

#### Abstract

Collaborative software development using the open-source model has become increasingly popular in bioinformatics; one prominent example of a successful project that adheres to this philosophy is BioMart ([www.biomart.org](http://www.biomart.org)). The BioMart project was initiated to address the challenges facing collaborating groups based in different geographical locations who wish to have unified access to their distributed data sources. BioMart has addressed these challenges by developing free, open source data federation software. BioMart's ability to quickly deploy a website hosting any type of data, user-friendly graphical user interface and several programmatic interfaces contributed to its success and adoption by many different types of projects around the world as their data management platform.

**ALL ARE CORDIALLY INVITED**