

## **Panel: Innovation on Digital Curation**

### **How can text mining scale to meet diverse and precise curation needs?**

Jo McEntyre, EMBL-EBI

#### **Abstract**

Expert curation is critical to data quality and reuse in the life sciences. Without question, curation accelerates access to research results, synthesizing data from thousands of labs into records that can be digested straightforwardly by humans and machines alike. Typically, curation involves filtering through masses of potentially useful research articles, identifying the most critical ones and then extracting key information, usually placing this into the context of a knowledgebase. We are interested in exploring how text mining can support such curation efforts, and furthermore, how we can take a more infrastructural approach to this challenge. The typical working relationship between a group of curators with specific interests and a text mining group that strives to support those precise needs will not scale as we strive to meet the increasing demand for quality curation. Furthermore, a siloed way of working means that we may miss collaborative curation opportunities. At Europe PMC, we are exploring how to combine text-mining, open access content, indicators of scientific quality and curator feedback on the Europe PMC platform. The goal is to engage the text mining community and curators to contribute open annotations, which, in the context of the rich metadata available in Europe PMC, may support article prioritization systems, or scan-reading tools for ascertaining the quality of individual papers (e.g. via the SciLite application). A new APIs for retrieving all annotations will shortly be released to encourage distribution and reuse in other contexts. Feedback on text-mined entities and relationships, with a view to improving algorithms is also desirable, from curators or other knowledgeable readers. Finally, to improve integration between the literature and data resources, we have developed a deep linking mechanism that enables precise linking between curated articles and data resources.