

## **Making Sense of Thousands of Messages**

### **March 2012 Update**

This project started in January. During the first part of the year, we worked on optimizing the text clustering application (named “Consilience”) to allow users to quickly navigate through possible clustering solutions, and choose their preferred clustering or partition for a given document set.

The IQSS and the Library team met in February to review the format of the email messages and differentiate between the actual text to be analyzed and the metadata fields. The Library team identified a first set of emails to work on, which includes around 500 email messages.

The IQSS team is now working on re-formatting this first set and extracting the metadata. In the next weeks, the re-formatted document set will be processed through the following steps: 1) calculate the word count in each email document, 2) run more than 100 clustering algorithms for the set, and 3) calculate new meaningful clustering solutions in the Bell’s space.