

Project SynFo

Research data management takes place under complex organisational framework conditions. Project *SynFo – Synergy Creation on the Operational Level of Research Data Management* aims to develop a researcher-centric solution for a pragmatic research data management. Based on an evaluation of concrete organisational and technical implementations in various institutions

involved in a research network, the commonalities are revealed in structural and legal context. The aim is to define organisational and technical interfaces and, where possible, to harmonise data management for researchers. A survey on the handling of digital research data conducted in the project forms the base of this analysis.

Interdisciplinarity, disciplinary Differences and temporal Changes

An Analysis of Research Data Management Survey Responses

Research is becoming more and more driven by digitized data and options as well as obstacles to handle it. The combination of an increasing proportion of transdisciplinary research and the often-formulated requirements for structured data management thus creates a difficult situation for researchers and providers of data management infrastructures. To get to know where the greatest differences and, hence, areas of harmonization are, we used the results of two surveys on the handling of research data: one from 2014 and a new one from 2018. We use R to analyse and visualise similarities and differences among survey participants' responses to get a better in-depth understanding of the survey results.

Analyses

Similarity of survey participants' responses

We calculated an adjacency matrix of Jaccard indices of all survey participants' responses compared with each other, set its diagonal and all other values but the maximum of each row to 0, created a graph and a visualisation from it.

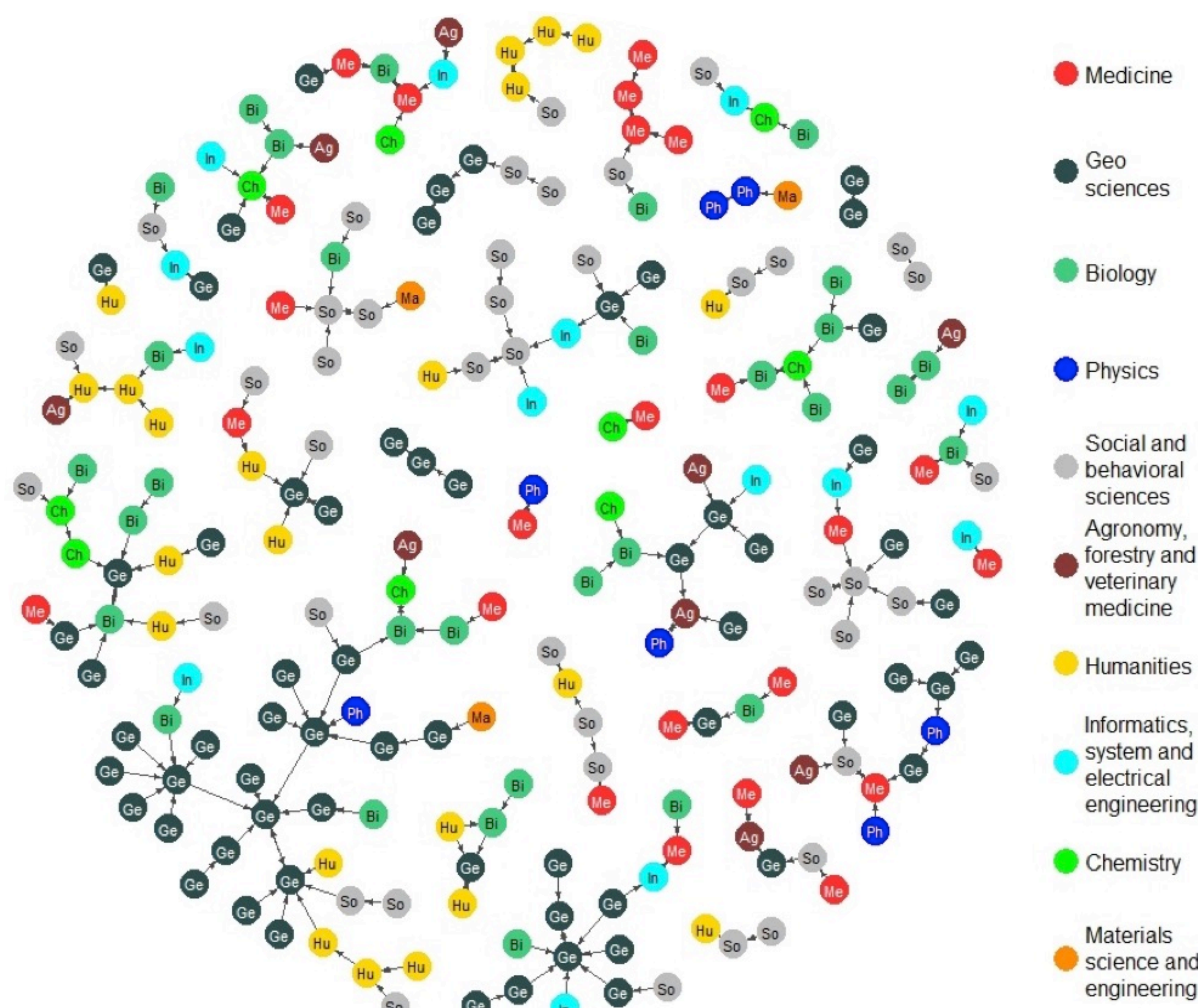
Differences of survey participants' responses

We apply principal components analysis (PCA) to study the answers of survey participants in 2018 and apply the model to the 2014 survey data.

Differences of survey participants' response patterns

We use Fisher's exact test for statistical significance to calculate p-values for each combination of the most frequently chosen survey items and create an adjacency matrix from the results. If the statistical significance (p-value) of the test between two survey items was 0.05 or less, we set the corresponding value in our adjacency matrix to 1 else we set it to 0. Therefore, our adjacency matrix represents relationships between survey item choices. We created a graph and a visualization from the adjacency matrix.

Maximum Jaccard-Similarity between Survey Participants' Responses



Findings

Similarity of survey participants' responses

The calculated matrix of Jaccard indices shows that researchers from the different areas of research answered most of the survey questions similarly. We suppose that this is related to interdisciplinary research in the four research foci at Kiel University: Kiel Marine Science, Kiel Life Science, Societal and Cultural Change and Kiel Nano Surface & Interface Science.

Differences of survey participants' responses

Regarding the survey responses' projections to the first two principal components for each area of research and the matrix of the PCA factor loadings, we find differences between survey participants' responses of humanities and medicine as well as between geo sciences and social and behavioural sciences mainly concerning the question, where the participants prefer to archive their research data. This seems to reflect that researchers of these areas of re-search are facing different ethical and legal challenges, and scientific cultures.

Differences of survey participants' response patterns

Compared to the survey in 2014, in 2018 the responses show that the need for advice on legal issues is not related to the need for storage and responsibilities anymore. We interpret this as a consequence of the imminent General Data Protection Regulation (GDPR) in the EU.