

Nachrichten aus dem Datenraum

Daten und Foresight – eine spannende Verbindung

In dieser Rubrik haben wir im Laufe des Jahres verschiedene Fragestellungen und Ansätze skizziert, mit denen Erkenntnisse aus Publikations- oder auch Social Media Daten gezogen werden können. Dabei wurde naturgemäß nur ein kleiner Teil dessen betrachtet, was möglich ist.

Die Verfügbarkeit umfangreicher Datentöpfe nährt bei vielen Entscheidungsträger*innen die Hoffnung, diese dafür einzusetzen, strategische Entscheidungen besser, schneller und vielleicht sogar automatisiert zu treffen. Am Fraunhofer INT untersuchen wir bereits seit einigen Jahren, wie wir Daten dafür nutzen können, Aussagen über technologische Zukünfte zu generieren oder zumindest zu untermauern. Dieses Forschungsfeld bezeichnen wir als Data Driven Foresight und das [KATI Projekt](#) ist ein wichtiger Bestandteil davon.

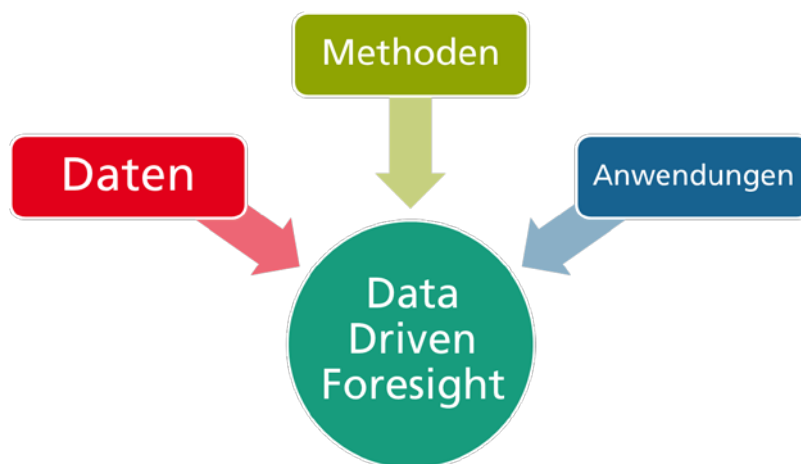


Abbildung 1: Die drei Dimensionen von Data Driven Foresight

Im Kern stützt sich Data Driven Foresight auf drei Säulen. Zuvorderst bilden natürlich Daten die Grundlage für alles Weitere. Deren Verfügbarkeit hat in den letzten Jahren stetig zugenommen. Neben Publikationsdaten, die die zentrale Zutat für das KATI-System bilden, sind dies Patentdaten, ökonomische Daten, aber auch Twitter-Daten (siehe den Beitrag vom Juni 2020 [Wie hältst Du's mit Technologien?](#)). Prinzipiell können viele verschiedene Datenquellen für einen Foresight-Prozess genutzt werden. Um dies zu erreichen, müssen sie verfügbar gemacht werden. Hierfür ist die Entwicklung eines geeigneten Datenmodells essentiell, damit man im weiteren Verlauf auf die Daten zugreifen kann. KATIs Datenmodell haben wir vor einiger Zeit in einem Youtube-Video skizziert, das man sich [hier ansehen kann](#).

Nachdem die Daten in geeigneter Art und Weise verfügbar gemacht wurden, können sie mittels datenanalytischer Verfahren erschlossen werden. Hier steht ein breites Portfolio an sehr unterschiedlichen Methoden zur Verfügung, vom klassischen Data Mining, über Methoden der Netzwerkanalyse hin zu Verfahren der Computerlinguistik und des Maschinellen Lernens. Diese zielen alle darauf ab, neue Einblicke in essentielle Aspekte eines Foresight-Prozesses zu finden.

Welche Methoden konkret zum Einsatz kommen, das hängt nicht nur von den verfügbaren Daten ab, sondern auch von den konkreten Fragestellungen oder Use Cases ab. Diese stellen den Anknüpfungspunkt an einen Foresight-Prozess dar, da sie letztlich bestimmen, welche

Daten und welche Methoden genutzt werden. Soll beispielsweise ein thematischer Überblick über die Forschung zu einem spezifischen Thema generiert werden, so liegt es nahe, wissenschaftliche Publikationen als Datenquelle zu nutzen. Um Teilthemen zu identifizieren, könnte man die semantische Ähnlichkeit zwischen den Publikationen messen und anschließend Gruppen von ähnlichen Publikationen bilden. In einem letzten Schritt müssen die so berechneten Ergebnisse noch visualisiert werden, wofür sich Verfahren aus dem Bereich der Netzwerkvisualisierung anbieten. Dieser Schritt ist von großer Bedeutung für den gesamten Prozess, da erst geeignete Visualisierungen es den Analyst*innen erlauben, die Daten explorativ zu erkunden und somit Informationen und Wissen zu generieren, welches dann im Foresight-Prozess genutzt werden kann.