

## **CITIESData: a smart city data management framework**

Das vorgestellte Paper beschreibt ein mögliches Konzept für die Datenverarbeitung von Daten, die in smart-cities anfallen. Smart Cities werden hierbei als Städte, welche eine große Auswahl an Informations und Kommunikationsdiensten nutzen definiert. Besonderes Augenmerk gilt hier den sogenannten Internet-of-Things(IoT) Geräten. Diese Geräte werden eingesetzt um bspw. Daten über Verkehrssituationen, Wetter, Lärmbelastung oder Luftqualität zu erheben. Im vorgestellten Framework geht es sowohl um das Sammeln und Aufbereiten, sowie das Anonymisieren und Veröffentlichen von Daten. Da die Daten aus unterschiedlichen Quellen gewonnen und somit auch unterschiedlicher Behandlung bedürfen, schlagen die Autoren vor eine Einteilung je nach Sensitivität der Daten vorzunehmen. Hierbei wird in sensitive, quasi-sensitive und öffentliche Daten unterschieden, für die jeweils unterschiedliche Verarbeitungsstrategien vorgeschlagen werden. Sensitive Daten werden als nicht veröffentlichbar angesehen, da sie vertrauliche Informationen enthalten und gewissen gesetzlichen Schutzmaßnahmen unterliegen. Die Einsicht dieser Daten soll also nur für autorisierte Zwecke, wie z.B. polizeiliche Ermittlungen, ermöglicht werden. Um die Privatsphäre bei der Verarbeitung der quasi-sensitiven Daten zu gewährleisten, werden im Paper einige Anonymisierungsansätze eingeführt. So beschreiben die Autoren das k-Anonymitätsmodell und geben zur Verbesserung dieses noch l-diversity und t-closeness an. Im Punkt der Datenaufbereitung ist Datenqualität für die Autoren der zentrale Faktor, daher schlagen sie ein Cleansingverfahren vor, welches regelbasiert die Qualität einzelner Daten quantifiziert und minderwertige Daten aus der Datenbank entfernt. Um die grobe Vorgehensweise des Frameworks zu skizzieren unterteilen die Autoren die Datenverarbeitung in mehrere Schichten. In der ersten Schicht geht es ausschließlich um die Datenakquise aus den unterschiedlichen IoT Geräten. In der Stagingschicht werden nun alle, noch heterogenen Daten, an einem Ort zusammengeführt. Im nächsten Schritt der Datentransformation, werden die Daten in ein einheitliches Format überführt, außerdem findet hier das Cleansing der Daten, sowie die Einteilung in die unterschiedlichen Sensitivitätsstufen statt. Non-sensitive Daten und die Metadaten von quasi-Sensitiven Daten können nun veröffentlicht werden, wogegen die restlichen Daten einen Anonymisierungsprozess durchlaufen, bevor sie veröffentlicht werden können. Im weiteren Verlauf des Papers wird ein detaillierterer Einblick in die Implementierung gegeben, sowie eine Beschreibung der Versuche, welche zur Evaluation des Frameworks genutzt wurden.

<https://link.springer.com/article/10.1007/s10115-017-1051-3>