

2023-09-20 Dokumentation Workshop „Datenplattformen“

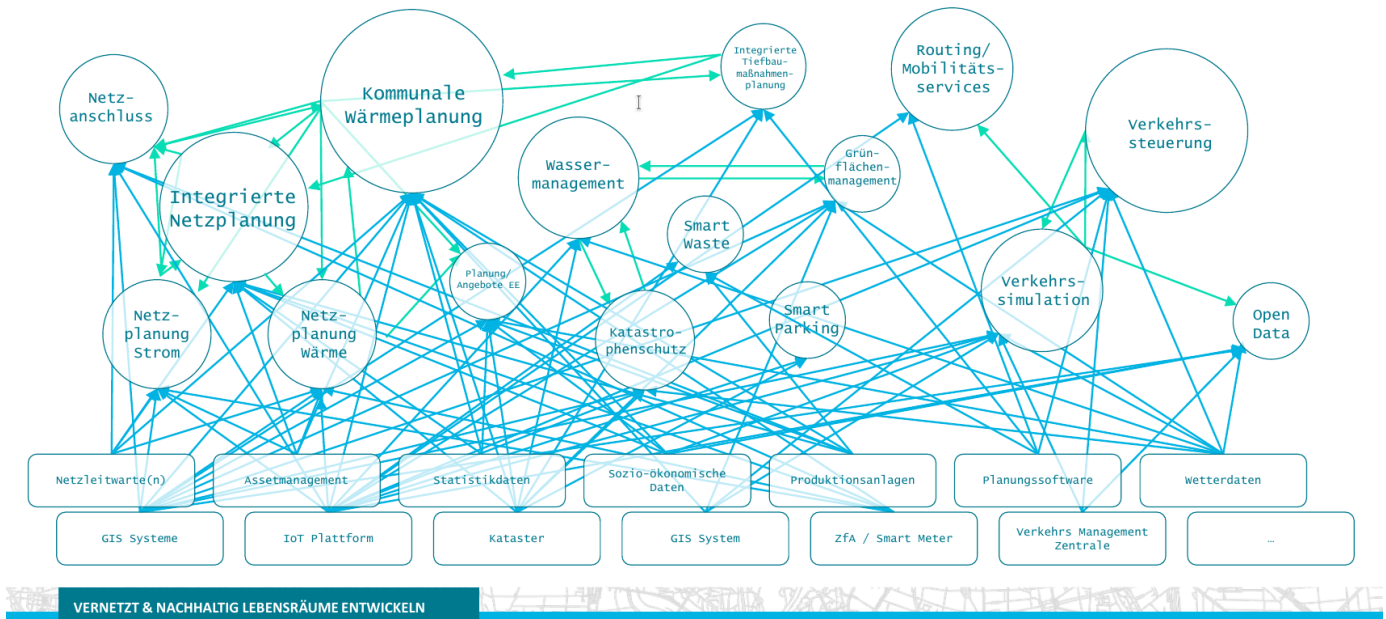


Workshopleiter:

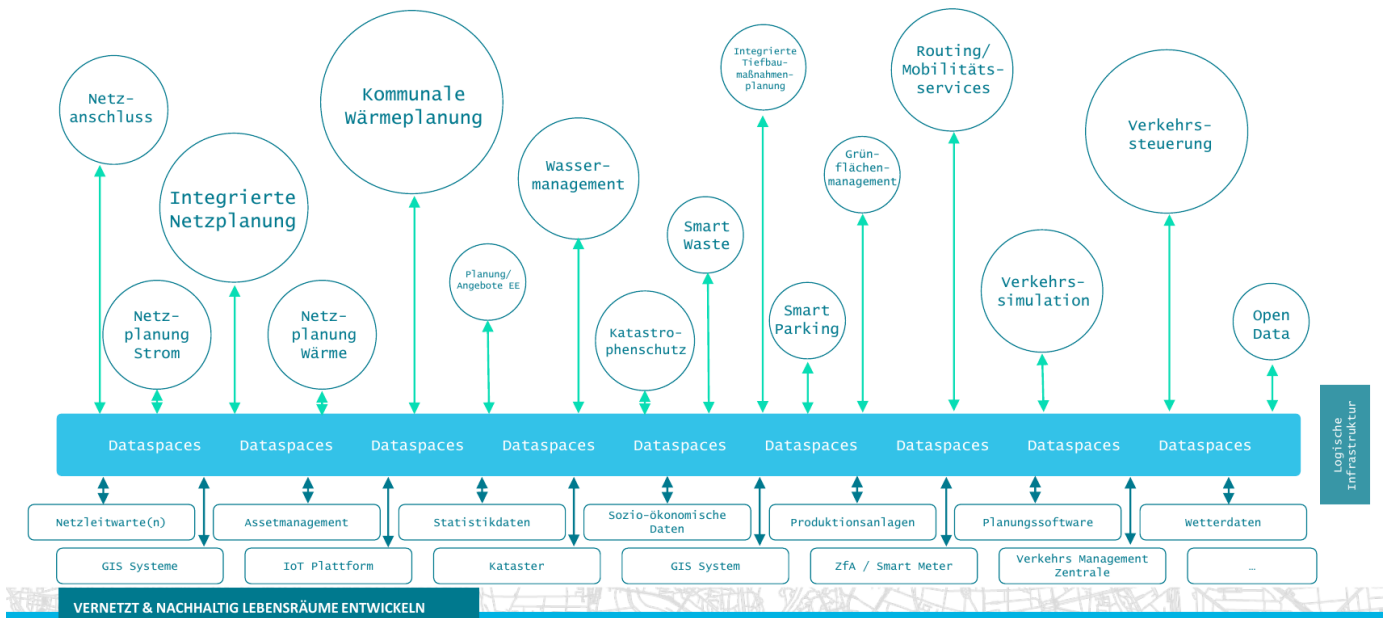
- Mathias Renner | Senior Consultant, IT-Architect, Project Manager; Frachtwerk GmbH / Civitas Connect e. V. -> LinkedIn Profil (<https://www.linkedin.com/in/mathias-renner-9b83a069/>)
- Marius Henkel | Sprecher der MPSC Themengruppe Datenplattform-Architekturen, Open Data Mensch und Smart City Teilprojekt-Leiter Stadtverwaltung Kaiserslautern -> LinkedIn Profil (<https://www.linkedin.com/in/marius-henkel-360661235>)

Intro

Civitas Connect unterstützt Mitglieder bei der Umsetzung der Digitalen Daseinsvorsorge. Die Kontexte bzw. Anwendungsfälle und den dafür nötigen Datenquellen sind zahlreich. Deren Verknüpfungen auch. Dadurch wird es unübersichtlich:

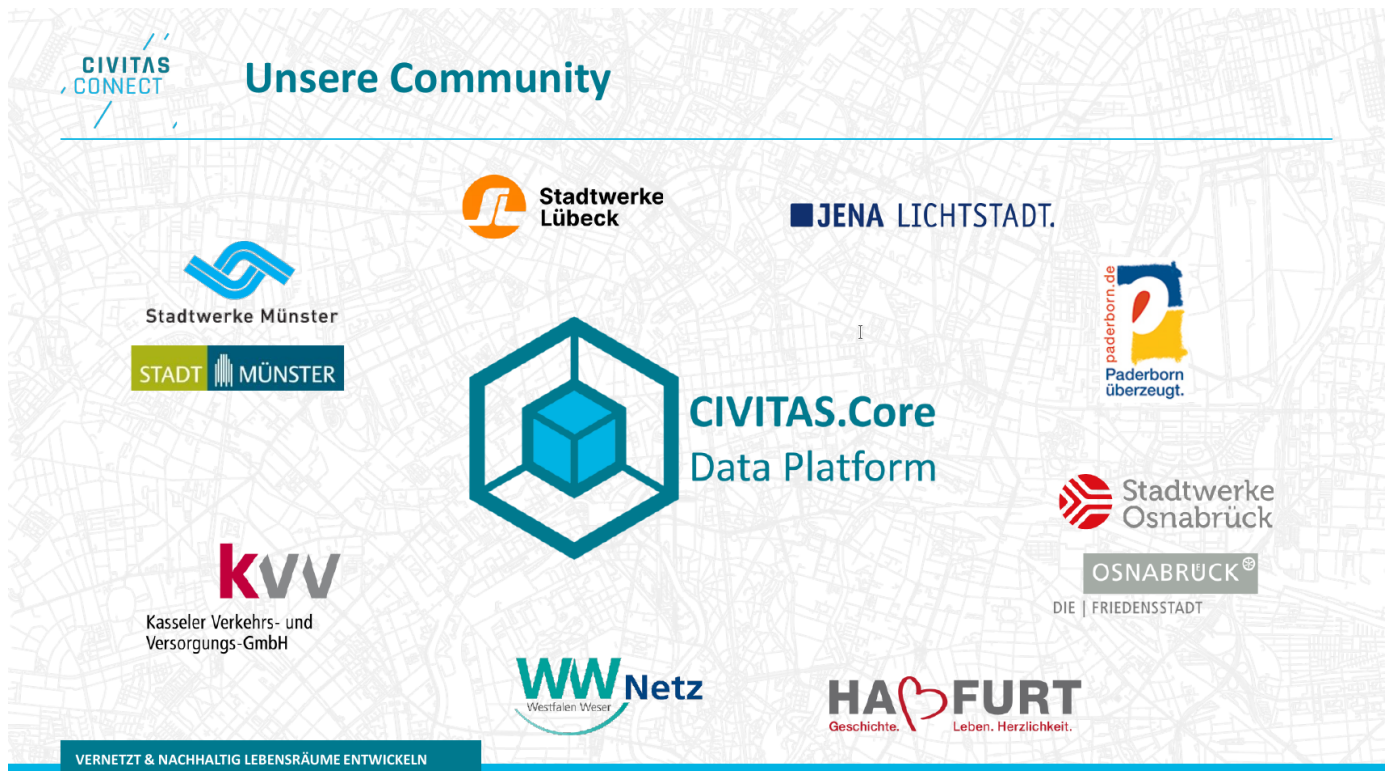


Lösungsvorschlag: Struktur durch Gruppierung der für die Kontexte/Anwendungsfälle nötigen Daten in “Dataspaces” bzw. “Datenräume”. Diese logische Schicht könnte von einer urbanen Datenplattform umgesetzt werden:



Breaking News: Civitas Connect hat vor 2 Wochen den Kick-Off durchgeführt für die Entwicklung einer Datenplattform-Technologie. Die Entwicklung wird in einer Gemeinschaft von Vereinsmitgliedern durchgeführt – als ausschließlich kommunale Kooperation. Diese Struktur ist in Deutschland einmalig!

Beteiligte:



Teil 1: Begriffsbestimmungen

Ziel: In Gesprächen zu Datenplattformen existieren unterschiedliche Definitionen. Wir möchten wesentliche Begriffe, wie UDP / IoT-Plattform, Datenportal / Open Data Portal, Digitaler Zwilling oder GDI mit Euch definieren und deren Relevanz in einer generischen Plattformarchitektur verorten, um das gegenseitige Verständnis zu erhöhen und so eine Grundlage für ein gemeinsames Glossar zu bieten.

Phase 1: Welche Dienste sollte ein digitales Ökosystem bieten?

Die Teilnehmer wurden in fünf thematische Gruppen aufgeteilt und haben Dienste innerhalb eines digitalen Ökosystems anhand der nachfolgenden abstrakten Architektur verortet. Diese Architektur orientiert sich im weitesten Sinne an der DIN SPEC 91357:



Am Boden war ein großer Kreis abgeklebt, mit den 5 Themengebieten an den Rändern. Dienste, die nahe am Nutzer sind sollten am äußeren Rand platziert werden. Dienste, die nahe an Sensoren sind als Datenquellen, eher in der Mitte:

- Mapserver/Geoserver/QGIS für Datenbereitstellung
 - Mapserver wird in Münster verwendet
- Geodatenportale/Metadatenkataloge wie Masterportal, Leaflet etc.
 - Zielgruppe ist intern als auch extern

Gruppe "Verwaltungsrevoluzzer":

- Meldesystem
- Verkehrsleitsystem
- Serviceportal
- Beteiligungsplattform für Anliegen von Bürgern an die Verwaltung

Gruppe "Open Data/Datenschutz":

- Herausforderung: Kollegen aus städtischem Amt haben oft Bedarfe an Daten, z. B. Gebäudebestand, die sie in einer strukturierten Form bereitstellen möchten innerhalb der Kommune als auch für andere Kommunen.
 - Dabei Klassifizierung von Daten wichtig (datenschutzrelevant, open data, ...): Wer macht das? -> "Data Governance"!
 - Zugriffskonfiguration auf Daten ist sehr wichtig, damit man steuern kann, wer auf Daten zugreifen darf!
- bei kommunaler Wärmeplanung ist Datenbereitstellung wichtig

Gruppe "IT-Experten":

- Rechenzentrum/Cloud/Server
- Data Lake als Datenpool, z. B. Azure Data Lake, als Sammelbecken für v.a. unstrukturierte Daten
- Datenpipelines mit z. B. Apache Kafka
- Streaming für Verarbeitung von Echtzeitdaten
- APIs
- IT-Security: Zugriffskontrolle, Authentifizierung, Identitätsmgmt, Autorisierung
- Künstliche Intelligenz mit z. B. chatgpt
- Apps: Dashboards (Apache Superset, Grafana), Smart City Cockpit

Gruppe "IoT-Enthusiasten":

- Hardware
- Middleware: = Datenplattform. Muss technologieoffen sein, viele Schnittstelle.
- Software: Endanwendung = Frontends Ist bei offener Datenplattform möglich.
- Datenkatalog: Nachschlagewerk für Daten: wie komme ich an die Daten ran?

Begriffsdefinitionen

Was für ein Verständnis haben wir für:

- GDI:
 - GIS: Geodaten sammeln und manipulieren
 - GDI: eher Rohdaten.
 - Geodatenportale sind an der Schnittstelle zwischen GDI und Nutzer
- Digitaler Zwilling
 - digitale Repräsentanz von Dingen oder Daten in der echten Welt
 - = jedes Werkzeug, was hilft, mit Daten zu arbeiten
 - z. B. Daten von Patient -> Patientenakte ist digitaler Zwilling
 - kann Darstellung und Simulation umfassen, die Daten verschneiden aus verschiedenen Bereichen, um damit dann Entscheidungsfindungen zu unterstützen
 - Es gibt oft mehrere Digitale Zwillinge, die dasselbe Thema aus unterschiedlichen Perspektiven betrachten. Jeder Digitale Zwilling hat bestenfalls immer einen Zweck/Anwendungsfall.
- (urbane) Datenplattform
 - alle Daten fließen rein und werden strukturiert
 - wesentliche Features:
 - übergreifendes Rechtmgmt für alle Daten
 - ETL-Prozess - Datentransformation/Datenaufbereitungen
 - Köln hat mehrere verteilte Datenbanken. Datenplattform ist ein "loses Gewebe", was ermöglicht, mit Data Governance Daten gut zu finden. Persistenz von Daten innerhalb der Plattform nur bei Bedarf.
- Digitales Ökosystem
 - Gesamtmenge aller technische und nicht technische (inkl. Datenstrategie etc.) Teilsysteme
- Dateninfrastruktur
 - uneinig. Entweder: Synonym zu digitalem Ökosystem oder nur der Kern des digitalen Ökosystems (insb. Sensornetzwerke).
- Datenraum/-Satz/-Punkt
 - Raum: Kontext von Daten, z. B. Luftqualität
 - Es kann ggf. auch Daten-Unterräume geben: "Luftqualität" kann ein Unterraum von "Umweltsensorik" sein etc.
 - Datensatz: Alle Daten von einem Sensore
 - Datenpunkt: Eine Messung / ein Messwert
- IoT Datenplattform
 - Verwaltung von IoT-Daten
 - Administration von Sensoren / Device-Mgmt
 - Berechtigungsmgmt

- Daten aus versch. Funktechnologien sammeln
- Verortung? IoT Plattform ist vorgelagert vor bzw. unterhalb der Daten-Plattform, selten Teil der (urbanen) Datenplattform
 - Manchmal auch gewachsenes System: Viele beginnen mit IoT-Daten und ergänzen dann später oben darüber eine urbana Datenplattform
- Datenmodell
 - umfasst Entitäten + Transformationen, und semantische Beschreibungen
 - Bsp: Business Objects for Energy gibt Modellstandards for für den Energiebereich. Die Vorlagen dort funktionieren aber nicht immer, am Ende ist es der Fachbereich, der das Datenmodell designed! Datenmodell müssen bei Bedarf frei modelliert werden können!
- Konnektor
 - Kommunikation zwischen Komponenten: z. B. zwischen IoT Plattform und Datenplattform, oder externen Datenquellen.
 - Ggf. passieren dabei Datentransformationen
- Open Source
 - ermöglicht Weiterentwicklung
 - ermöglicht & erhöht Datensouveränität
 - sollte bei Finanzierung von Steuergeldern der Standard sein, weil man so Ergebnisse an die Gesellschaft zurückgeben kann.

Hinweis: Es gibt eine neue DIN SPEC, die in Q1/2024 released wird, die ähnliche Begriffsdefinitionen umfasst.

Teil 2: Erfolgsfaktoren: Was sind Erfolgsfaktoren für die Nutzung einer urbanen Datenplattform?

Ziel: Was ist nötig für die effiziente Nutzung einer urbanen Datenplattform (UDP)? Wie können Daten effizient gemanaged werden? Welche Benutzeroberflächen brauche ich für die Erstellung neuer Datensätze und Use Cases?



- Kommunikation und Zusammenhalt:
 - Alle Abteilungen müssen miteinander kommunizieren.
- Dokumentation:
 - Ergebnisse mit anderen teilen.
 - Guidelines: z.B. Namenskonventionen, How-Tos etc.
- Vertrauen:
 - Menschen möchten Daten nicht rausgeben.
 - Vertrauen dafür ist wichtig.
 - Beispiele/Präzedenzanwendungen können helfen. evtl. auch per Rapid Prototyping zeigen, wie etwas aussehen könnte.
 - Problem: Bei Datenplattform schwierig, weil Nutzen nur mittelbar. Gefahr besteht, dass man sich zu sehr auf die Umsetzung einzelner, gut darstellbarer Quick-Wins fokussiert und die generische Infrastruktur nicht umsetzt.
 - Besser: Kompetenzen aufbauen, damit Menschen das Konzept besser verstehen.
- Zentralisierter Ansatz / Dienstleistung:
 - Personal in den Kommunen für Betrieb und die Weiterentwicklung einer eigenen Datenplattform meist kaum vorhanden oder knapp
 - Mandantenfähigkeit schafft effizientere Ressourcennutzung. Viele Use Cases in Kommunen sind identisch.
- Sicherheit:
 - Teilweise kritische Infrastruktur. Zielkonflikt zwischen Sicherheit und Umsetzungs-Geschwindigkeit/Flexibilität
 - Auf Manipulationssicherheit sollte geachtet werden, z.B. bei Sensorgestützter Schaltung von Systemen. Sensorik kann manipuliert/getauscht werden.
- Skalierbarkeit/Performanz:
 - Muss für alle User gut und reibungslos funktionieren.
- Modularität:
 - Für bessere Skalierbarkeit, Kosten und Aufwand besser zuschneiden zu können. Umfasst sowohl Funktionen als auch Performance.
- Datenintegrität:
 - Rohdaten sinnbehaftet abspeichern / mit kontextinformationen anreichern.
- Datenkatalog:
 - Was existiert an Daten? Schnelles Verständnis, was diese Beinhalten.
 - Erfolgsfaktor: Aktualität
- Offene Schnittstellen:
 - Standard-Schnittstellen. Wohldokumentierte REST-Schnittstelle. Semantische Schnittstellenbeschreibung.
 - Erkenntnis: NGS-LD spielt dabei bislang keine große Rolle
 - Input Ulrich Ahle:

- NGSI-LD ist ETSI Standard. Beantwortet die Frage: “Wie übertrage ich?”
- Smart Data Models beantworten die Frage “Welche Daten nutze ich?”
 - Anderer, eher Geodaten-Fokussierter Standard: SensorThingsAPI (OGC-Standard)
- Datennutzbarkeit erhöhen, Daten verschneiden
- Open Source:
 - notwendig für Übertragbarkeit von Lösungen. Veröffentlichung auf opencode.de (<http://opencode.de>) für Förderprojekte obligatorisch
 - Für Konnektoren besonders wichtig
- Einfache Benutzbarkeit:
 - No Code, Low Code. Es soll wenig Spezialwissen erforderlich sein, um Dinge umsetzen zu können.
- Gute Frontends

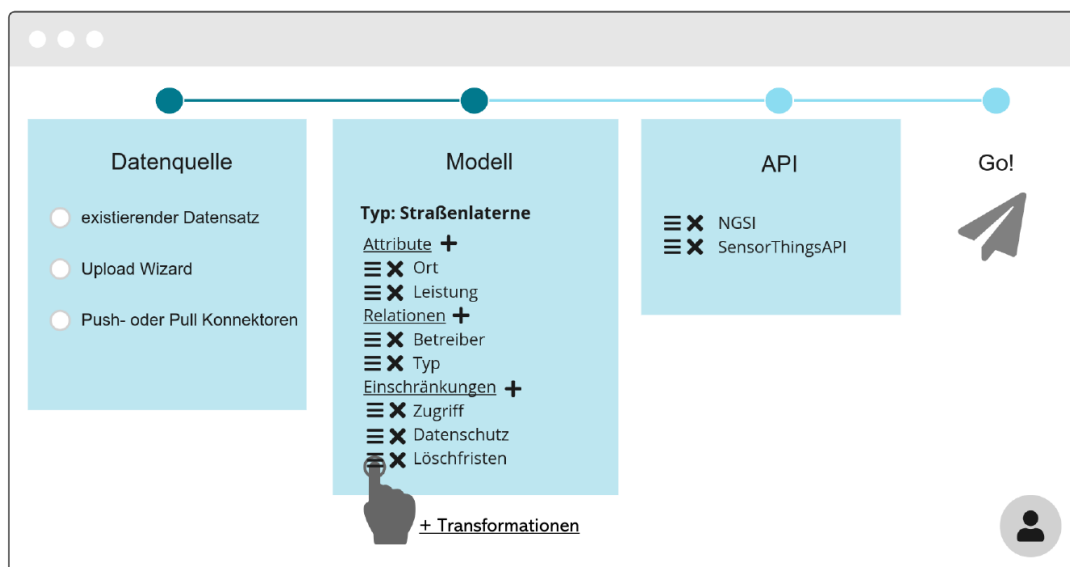
Kann ein “Datenmodell-zentrierter Ansatz” ein Erfolgsfaktor sein?

Wir stellen als Diskussionsgrundlage hierfür einen Ansatz vor, der die freie Modellierung von Datenmodellen in den Mittelpunkt stellt.

Eine Benutzeroberfläche könnte wie folgt aussehen:



Workflow: Modell-zentrierter Ansatz.

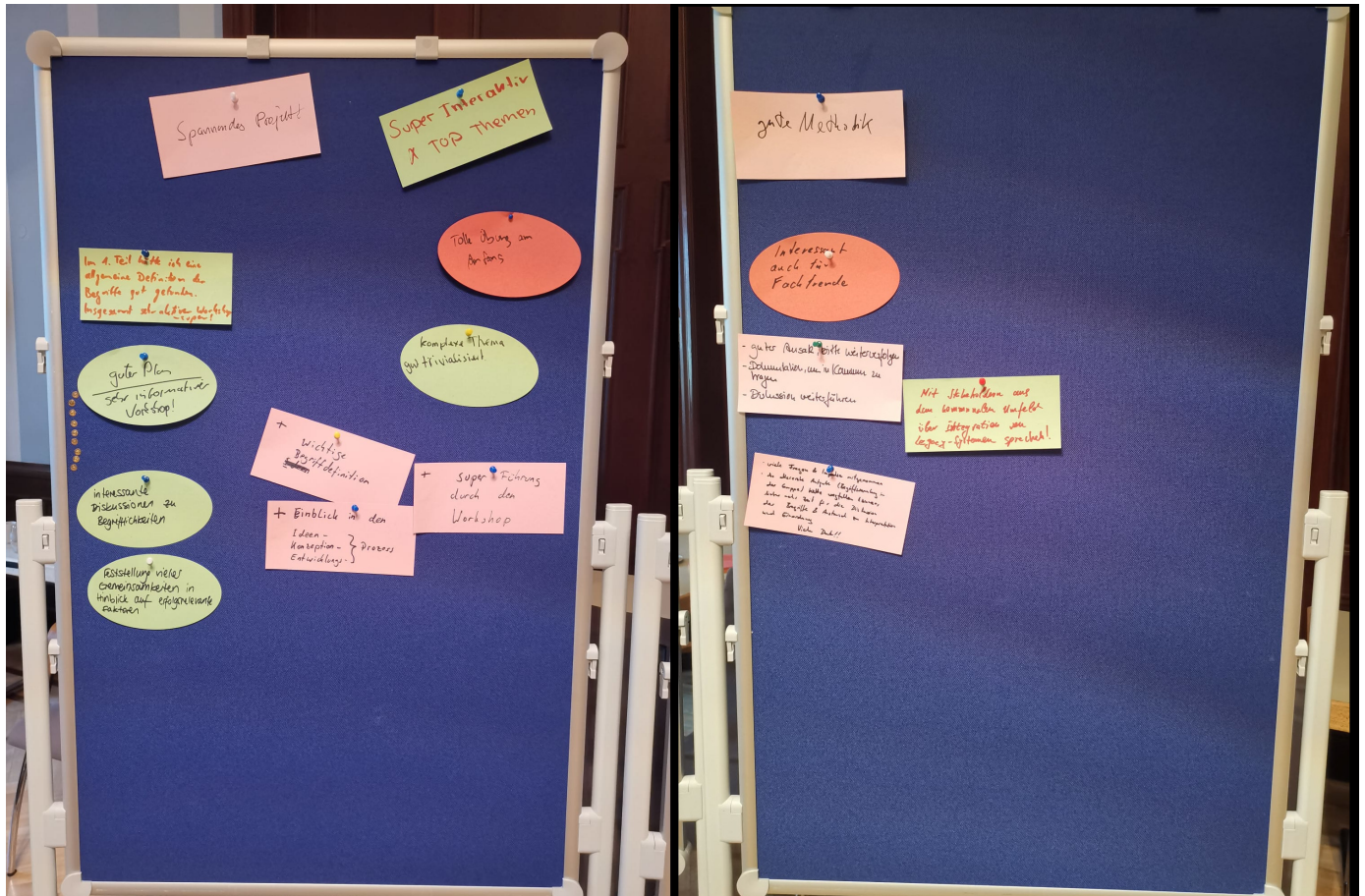


8

Idee: Nutzer der Datenplattform erhalten eine Benutzeroberfläche als Self-Service Portal. Damit können sie selbstständig Daten in die Plattform laden, deren Modell erstellen inkl. Zugriffe & Transformationen (z. B. aufgrund von Datenschutzanforderungen) und entscheiden, über welche Schnittstelle die Daten bereitgestellt werden sollen.

Civitas Connect überlegt in der Plattfromtentwicklung so etwas zu entwickeln. Jegliches Feedback ist willkommen!

Feedback & Outro



Mathias Renner
Consultant & IT-Architekt Frachtwerk
mathias.renner@frachtwerk.de



Marius Henkel
Stadtverwaltung Kaiserslautern
marius.henkel@kaiserslautern.de

Wir liefern Bausteine für die digitale Infrastruktur, die kommunale Akteure befähigt, ihre Daten selbstständig zu verwalten und deren Potenzial nachhaltig auszuschöpfen.



Danke!

VERNETZT & NACHHALTIG LEBENSÄUERE
ENTWICKELN