

## **Zusammenfassung vom Workshop *Herausforderungen des Multistakeholderismus im steirischen Datenökosystem* (25.03.2021)**

Am 25.03.2021 trafen sich über 20 Stakeholder aus der Steiermark zum IDE@S Workshop zum Thema *Herausforderungen des Multistakeholderismus im steirischen Datenökosystem*. Ziel des Workshops war es, die vorhandenen Ressourcen sowie auch Wünsche hinsichtlich der technischen Infrastruktur für eine kollaborative Datenumgebung zu erheben, um sich im weiteren Verlauf Gedanken zu Schnittstellen machen zu können. Die Teilnehmenden aus Wissenschaft, Wirtschaft sowie öffentlicher Verwaltung anwesend von *TU Graz, MedUni Graz, Karl-Franzens-Universität Graz, Universität Leoben, FH Joanneum Kapfenberg, Joanneum Research, Biobank, KAGes, fragmentiX, Innophore, Invenium Data Insights* und *Kapsch* hatten dabei im Rahmen des Workshops die Gelegenheit sich zu folgenden Themengebieten in Diskussionsrunden auszutauschen:

- (1) Vorhandene Kooperationsprojekte – beteiligte Organisationsformen, Cloudplattformen und kooperative Tools sowie Daten & Standards im individuellen Bereich*
- (2) Vorhandene und gewünschte Datenlösungen mit Fokus Middleware*
- (3) Individueller Anschluss an Kooperationsnetzwerk und generelle Herausforderungen*

Als Ausgangspunkt wird ein System verteilter Datenkontrolle betrachtet, welches Datensouveränität für Dateneigentümer\*innen gewährleisten muss. Dieses System umfasst die Speicherung generierter Daten von verschiedenen Organisationen auf deren bzw. geteilter Infrastruktur.

**Während der Diskussion haben sich folgende essentielle Kernpunkte herauskristallisiert, die bei der Entwicklung einer gemeinsamen Infrastruktur bearbeitet werden müssen: IT-Sicherheit, Datenschutz, und Standards für Daten, sowie für Prozesse.**

Die Erhebung der bestehenden technischen Infrastruktur zeigt bei diversen Stakeholdern auf, dass die lokale Speicherinfrastruktur in der Regel ausreicht, jedoch relativ schnell Grenzen an Rechenleistung (im Sinne von zeitlich abhängigen Gleitkommaoperationen) der eigenen Struktur erreicht werden. Parallel erscheinen Cloudanwendungen an Bedeutung zu gewinnen.

Die Erhebung von Werkzeugen für kooperatives Arbeiten ergab ein äußerst vielfältiges Bild: Eigenlösungen und lokale Server sowie Secure Clouds, aber auch Dienstleistungsfirmen mit kommerziellen Produkten wie Citrix, Clouds wie Azure, AWS, oder auch Spark, SAP (HANA), Oracle, Datenlösungen heterogener Struktur, individuelle Laborsysteme, und andere Datenbanken, aber auch verteilte Datei-Systeme. Die genannten Beispiele waren zahlreich und reichten bis hin zu Email-Anhängen einzelner Kooperationspartner. Geeignete Schnittstellen

nach außen fehlen meist und das Bewusstsein für ausreichende Datensicherheit muss in vielen Fällen bei Kooperationspartnern noch erhöht werden.

Eine weitere Perspektive ergibt sich durch die Frage nach Datenschutz-Leitlinien. Ein Mangel an Leitlinien führt oftmals dazu, dass Kooperationen entweder gar nicht oder nicht in adäquater Weise durchgeführt werden können. Um dies zu verhindern, ist eine Klärung von Verantwortlichkeiten, insbesondere bei Fehltritten, vorab von Bedeutung. Spezifische Datenaufbereitungen sind je nach individuellem Bereich für eine Verfügbarmachung und Datenfreigabe von Nöten. Dies schließt eine Datenklassifizierung mit ein.

Bezüglich vorhandener Datenstandards wurde wiederholt das Format CSV (comma-separated values) genannt. In vielen Bereichen (z.B. Hochleistungsrechnern) sind Eigenlösungen in

Shell-Skripts noch immer Usus und führen durch komplexe Workarounds auf lange Sicht zu Schwierigkeiten bei der Instandhaltung bzw. Wartung. Eine Etablierung gewisser Standards ist daher jedenfalls erforderlich. Daraus ergibt sich ein weiterer Kernpunkt aus den gegebenen Fragestellungen, und zwar die Standardisierung. Große Konzerne verwalten ihre Daten individuell und in proprietären Formaten, abseits von anerkannten Standards. Gewünschte Daten-Lösungen sollen hingegen offene Datenstandards und deren Etablierung umfassen. Letzterer Schritt könnte durch Organisations-übergreifende Entwicklungen (z.B. Datenexperten mit Domänenexperten) realisiert werden. Manche Bereiche (z.B. digitalisierter 3D-Guss, Text-basierte Kanzlei-Daten, biomedizinische Daten und aufwendig oder nur manuell zu anonymisierende klinische Daten) erfordern Austauschformate und eine Reprozessierung, die zur Datenweitergabe notwendig ist. Dieser Schritt ist essentiell, um eine nachhaltige Datenverwendung zu ermöglichen und alte Daten für neue Modelle zu nutzen.

Im Bereich Middleware sind bereits viele Konstrukte aus Open-Source Bausteinen konstruiert. Zusätzlich gewünschte und benötigte Funktionen umfassen: homomorphe Verschlüsselung, Anonymisierungsservices bzw. allgemeiner Marketplace für Services, Metadatenvorgabe, Single Sign-on oder auch eine mögliche Anbindung an Hochleistungsrechner (z.B. Vienna Scientific Cluster).

Es wurden auch einige Ideen für ein erfolgreiches Projektergebnis von IDE@S genannt. Unter anderem sollte das „Kernöl“, also die steirischen Vorteile des Projekts, betont werden, und damit auch „Entscheidungsträger\*innen“ angesprochen werden. Je nach Fortschritt des Projektes sollten gerade in anfänglichen Phasen Vorteile einer steirischen Plattform gegenüber den europäischen Großprojekten beworben werden. Ein Beispiel hierbei liegt bei der Verwendung von offenen Standards im Gegensatz zu geplanten Integrationen in Gaia-X. Weiters liegt gerade in der Steiermark

konzentrierte Vielfalt an IKT-Infrastruktur vor mit einer Vielzahl an Hochschulen, die gebündelt werden müssten. Der Datenfluss sollte über den IDE@S Hub geordnet werden und als „Wache“ im Sinne von neutralem Verwalter der Daten und zugehöriger Überwachungsfunktion fungieren. Hier sollte der Mehrwert hervorgehoben werden, dass sich Stakeholder keine Gedanken zur Sicherheit der Daten machen müssen. Zudem sollte ein möglicher Anschluss an größere Projekte gewährleistet sein und in vielleicht späteren Projektphasen nachrücken, sobald europäische Strukturen etabliert sind und anfangs erst die Benefits einer steirischen Plattform beworben werden.

Des Weiteren sollen Beschreibungen von realen aber auch fiktiven, zukünftigen, Use-Cases das Aufzeigen des praktischen Nutzens für mögliche Stakeholder erleichtern. Um neue Stakeholder zu gewinnen und bestehende zu halten, sollte ein System der Gegenleistungen geschaffen werden. Ein rein altruistisches System erscheint nicht machbar. Als mögliches Modell wurde der sogenannte „Styriahub“ mit „Styriatokens“ entworfen, eine Art digitales monetäres System, um den Tausch von Daten, Infrastruktur und Dienstleistungen zu ermöglichen und zu fördern. Für einen dezidierten Erfolg erfolgte auch ein Aufruf aus der Audienz zum demonstrativen Teilen der eigenen Kapazitäten, um einen Prototyp im kleineren Rahmen zur Veranschaulichung zu kreieren. All diese Aspekte können in das zu Beginn vorgestellte System verteilter Datenkontrolle (IDE@S Hub) einfließen und das Konzept konkretisieren. In den nächsten Workshops werden die identifizierten Schwerpunkte, insbesondere Standardisierung und Sicherheit, im Detail behandelt und Anforderungen erhoben, sowie technische Konzepte erarbeitet.

Wir bedanken uns bei allen Workshop-Teilnehmenden für ihre aktive Beteiligung und wertvollen Beiträge!

*Autorinnen: Claire Jean-Quartier, Sabrina Knopper, Sarah Stryeck*