

Visionspapier: Grüne Datenwirtschaft

Autor*innen: Dr. Sarah Stryeck, Dr. Claire Jean-Quartier, DI Mag. Günther Tschabuschnig

Nachhaltige Datenwirtschaft

Daten sind ein Rohstoff unserer Zeit. Rohstoffe werden üblicherweise abgebaut und in unveränderter Form direkt verwendet, oder prozessiert und in verarbeiteter Form konsumiert. Dasselbe passiert auch mit Daten. Daten werden erzeugt bzw. erhoben. Die aus Daten gewonnenen Informationen können direkt weiterverwendet werden, oder die Daten werden prozessiert, um die gewünschte Information zu erhalten. Diese Prozesse für den Abbau, die Prozessierung und Weiterverarbeitung von Rohstoffen sind in vielen Fällen ressourcenintensiv – dasselbe gilt auch für die Verarbeitung von Daten.

Viele europäische oder internationale Organisationen streben danach nachhaltiger zu agieren und setzen Ziele zur Ressourcenschonung. So definierten die vereinten Nationen 17 Ziele für nachhaltige Entwicklung (Sustainable Development Goals¹). Diese sollen der Sicherung der nachhaltigen Entwicklung auf ökonomischer, sozialer sowie ökologischer Ebene dienen. Der europäische Grüne Deal² wurde von der europäischen Union zur Bewältigung der klima- und umweltpolitischen Herausforderungen ins Leben gerufen. Der Maßnahmenplan soll die EU-Wirtschaft nachhaltiger, sauberer und ressourceneffizienter machen. Ziel ist es bis 2050 die Netto-Treibhausgasemission auf 0 zu reduzieren, bis 2030 soll eine Senkung um mindestens 55% im Vergleich zu 1990 erreicht werden.

Neben den großen Initiativen werden auch bereits konkrete Prozesse und Standards für nachhaltige Rohstoffnutzung definiert. So stellt beispielsweise das Renewable Energy Certificate System (RECS)³ ein standardisiertes Zertifikat als Nachweis für die Erzeugung von erneuerbarer Energie in Europa dar. Für Smart Buildings werden ökologisch unbedenkliche und emissionsarme Baustoffe genutzt. Diese werden mit dem Environmental Product Declarations (EPD)⁴ gekennzeichnet. Verantwortungsvolle Waldwirtschaft wird mit dem Forest Stewardship Council (FSC) Standard⁵ bzw. dem Programme for the Endorsement of Forest Certification Schemes (PEFC) Standard⁶ zertifiziert. Es gibt noch viele weitere Beispiele für die Zertifizierung

¹ <https://sdgs.un.org/goals>

² https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal_de

³ RECS - www.e-control.at

⁴ EPD Programm | IBU - Institut Bauen und Umwelt e.V. (ibu-epd.com)

⁵ [Verarbeitung / Handel \(fsc-deutschland.de\)](http://www.fsc-deutschland.de)

⁶ [PEFC Austria - PEFC Austria](http://www.pefc-austria.at)

von nachhaltigen Prozessen bzw. Produkten in Lebensmittelbranche, Baubranche und viele mehr.

Nachhaltigkeit ist auch auf Datenebene wichtig. Im Bereich Datenwirtschaft liegt das Hauptaugenmerk der Initiativen derzeit auf der gemeinsamen Nutzung von Daten und der Schaffung der dazu benötigten Infrastrukturen (z.B. GAIA-X⁷, European Open Science Cloud⁸). Die Qualität rückt immer mehr in den Vordergrund. Hier liegt der Fokus aber primär auf der Qualität der Datensätze (Vollständigkeit, Konsistenz, Aktualität und Konformität) bzw. der Dokumentation (z.B. ISO 9000, ISO 25000, ISO 19158, ISO 19157, ISO 25010, ISO 8000(-61)). In manchen Fällen wird die Eignung der Daten in einem bestimmten Anwendungsfall betrachtet.

Derzeit gibt es jedoch noch keine Zertifikate oder Standards für Daten, welche nachhaltig erhoben, prozessiert und wiederverwendet wurden. Blauer Engel entwickelte einen ersten Ansatz für Ressourcen- und energieeffiziente Softwareprodukte.⁹ Durch die steigende Verwendung von Daten und die Entwicklung von zahlreichen datengetriebenen Unternehmen ist es unerlässlich die Nachhaltigkeitsaspekte zu berücksichtigen. Dazu gehören Fragestellungen wie

- Woher kommen die Daten mit denen wir arbeiten?
- Wie klimaschädlich oder klimafreundlich wurden diese Daten erhoben?
- Ist die Sensorik für die Datenerhebung aus umweltverträglichen Materialien?
- Welchen Einfluss hat Speicherung oder Prozessierung der Daten auf das Klima?
- Welche Prozesse sollen nachhaltiger gestaltet werden?

Ziel dieser Publikation ist es die relevanten Aspekte für "grüne Daten" entlang des gesamten Datenlebenszyklus darzustellen. Es werden Fragestellungen definiert, die für einen Datensatz beantwortet werden müssen, um die Nachhaltigkeit zu spezifizieren. Diese Sammlung von relevanten Fragestellungen bildet eine Grundlage für die Entwicklung von Metriken und Standards zur Gestaltung eines Gütesiegels für grüne Daten.

Nachhaltigkeitsaspekte rund um den Datenlebenszyklus

Wie Rohstoffe durchlaufen auch Daten einen. Konkret werden die Daten generiert / erhoben, prozessiert, analysiert, aufbewahrt und schlussendlich (nach-)genutzt oder vernichtet. In all den Phasen dieses Datenlebenszyklus wird mit Ressourcen-verbrauchenden Technologien und

⁷ [GAIA-X - Home \(data-infrastructure.eu\)](https://gaia-x.eu/)

⁸ [Advancing Open Science in Europe | EOSC Association](https://www.eosc.eu/)

⁹ [Blauer Engel, Softwareprodukte, ressourcen- und energieeffizient, transparente Schnittstellen | Blauer Engel \(blauer-engel.de\)](https://www.blauer-engel.de/)

Infrastrukturen gearbeitet, um die Daten zu erheben, zu verändern oder zu verarbeiten. Schätzungen zufolge verbrauchen Rechenzentren über 200 Terawattstunden (TWh) pro Jahr; das ist mehr als der gesamte Energieverbrauch einiger Länder¹⁰. Zusätzlich steigt der Energieverbrauch durch Kryptowährungen weiter. Deren Leistung beläuft sich auf über 50 Gigawatt pro Tag¹¹, was der benötigten Leistung eines Jahres in Polen entspricht. Die Prognose des zukünftigen Energiebedarfs von datenverarbeitenden Systemen stammt von Anders Andrae (Huawei Technologies Schweden); er prognostiziert, dass sich der Stromverbrauch von Rechenzentren bis 2030 voraussichtlich auf 8 % des prognostizierten weltweiten Bedarfs erhöhen wird.¹² Im Folgenden werden diese Ressourcen-nutzenden Aspekte in den jeweiligen Phasen entlang des Datenlebenszyklus erläutert, sowie relevante Fragestellungen zur Bewertung der Nachhaltigkeit der Daten definiert.

Datengenerierung / -erhebung

Daten werden zunächst generiert oder erhoben. Das kann auf verschiedenste Arten erfolgen: durch Messinstrumente, durch Sensorik, aber auch durch die manuelle Erhebung (z. B. manueller Eintrag in Datenbank, Transkription von Interviews). Hier kann vor allem die Verwendung von nicht ressourceneffizienten Sensoren eine tragende Rolle für die Umwelt spielen. Es gibt bereits viele Ansätze, um umweltfreundlichere Sensoren zu entwickeln wie beispielsweise aus naturfaserverstärkten Biopolymeren¹³.

Der Großteil der generierten Daten stammt aus drei Hauptquellen: Daten von sozialen Medien, Maschinendaten (inkl. 5G) und Transaktionsdaten. Daten von sozialen Medien werden produziert durch die Erstellung von Likes, Tweets & Retweets, Kommentaren, Video-Uploads und allgemeinen Medien, die hochgeladen werden. Diese Art von Daten bietet unschätzbare Einblicke in das Verhalten von Kund*innen und ist essenziell für Marketinganalysen. Maschinendaten sind Informationen, die von Industrieanlagen und unterschiedlichsten Arten von Sensoren (z. B. Druck- oder Schallsensoren, Medizingeräte, Satelliten) stammen. Es wird erwartet, dass diese Art von Daten exponentiell anwächst¹⁴, da das Internet der Dinge (zu Englisch: Internet of Things (IoT)) immer allgegenwärtiger wird. Darüber hinaus wird das globale Datenwachstum in naher Zukunft weiter angekurbelt, da viele Länder den Ausbau von 5G-Netzen vorantreiben. So hat beispielsweise China Anfang November 2019 ein umfangreiches 5G-Netz für seine Telekommunikationsbetreiber*innen eingerichtet.¹⁵

¹⁰ <https://www.nature.com/articles/d41586-018-06610-y>

¹¹ <https://ccaf.io/cbeci/index>

¹² <https://doi.org/10.3390%2Fchalle6010117>

¹³ [Umwelt-freund-li-che Sen-so-rik aus dem Drucker \(apa.at\)](https://www.umweltfreundliche-sensorik.de/)

¹⁴ https://www.nikkoam.de/articles/2020/whats_causing_the_exponential

¹⁵ https://www.nikkoam.de/articles/2020/whats_causing_the_exponential

Transaktionsdaten werden aus allen täglichen Transaktionen generiert, die sowohl online als auch offline stattfinden. Rechnungen, Zahlungsaufträge, Lageraufzeichnungen, Lieferscheine – alle Daten werden als Transaktionsdaten bezeichnet. Auch bei der Erhebung bzw. Generierung von Daten können irrelevante Daten entfernt werden.

Relevante Fragestellungen

- Wie werden die Daten erzeugt?
- Werden im Falle von Messdaten energieeffiziente Sensoren verwendet?
- Welche Speicherressourcen werden für die Daten verwendet?
- Wie werden irrelevante Daten bei der Erhebung bzw. Generierung von Daten behandelt?

Datenprozessierung

Die Datenprozessierung ist jener Schritt, bei dem die Daten bereinigt und für Analysen vorbereitet werden oder Rohdaten in höherwertige Datenprodukte umgesetzt werden. Hier kann vor allem Datenbereinigung die Menge der gespeicherten Daten, durch die Entfernung von unnötigen Replikaten (z. B. durch Deduplikationssoftware), reduzieren. Zusätzlich kann die Verwendung von Speicher- und Rechenressourcen mit erneuerbaren Energien den ökologischen Fußabdruck um ein Vielfaches verbessern.

Für die Prozessierung von großen Datenmengen gibt es auch Ansätze die versuchen grüne Energie bevorzugt zu verwenden. Green Hadoop beispielsweise gleicht Energiebedarf und -verfügbarkeit ab, und setzt, wenn möglich, auf grüne Energie, um die Nutzung von brauner Energie zu minimieren.¹⁶

Server für Datenprozessierung sind im Durchschnitt nur 20-30 Prozent der Analysezeit beschäftigt. Energiemanagement, also die Nutzung des Ruhezustandes, hilft den Energieverbrauch zu verringern. Es erfordert daher Richtlinien, um inaktive Server nicht mit Strom zu versorgen.¹⁷ Neben der Energiemanagement gibt es noch eine Vielzahl von organisatorischen Maßnahmen¹⁸ zur Steigerung der Energieeffizienz wie Auslastungsoptimierung von Anlagen, Prozessharmonisierung und homogenisierte Serverlandschaft, Beschleunigung der Erstellung von computergestützten Dienstleistungen,

¹⁶ <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.453.3699&rep=rep1&type=pdf>

¹⁷ <https://www.hindawi.com/journals/tswj/2014/452863/>

¹⁸ [A Taxonomy and Survey of Green Data Centers | IEEE Conference Publication | IEEE Xplore](#)

Redundanzvermeidung, Energie-Überwachung und Metriken^{19,20,21} zur Bewertung der Effizienz der Umweltbemühungen.

Zusätzlich ist zu erwähnen, dass Server nur einen Teil des Energieverbrauchs im Serverraum verursachen. Vor allem die Kühlung von Serverräumen kann hier ein Viertel der Stromkosten ausmachen.²² Hier kann durch Verwendung von thermisch angetriebenen Kälteanlagen (Absorptionskältemaschine) der Energieverbrauch deutlich reduziert werden (Stichwort Solare Kühlung²³). Auch die Verwendung der Abwärme für die Beheizung anderer Gebäudebereiche ist eine Möglichkeit Ressourcen zu schonen.²⁴ Daneben sollte auch der Energieverbrauch der Speichertechnologien mitbedacht werden.

Relevante Fragestellungen

- Werden unnötige Datenmengen durch Datenbereinigungsverfahren entfernt?
- Welche Speicher- und Rechnertechnologien werden für die Verarbeitung und Prozessierung verwendet?
- Wird in allen verfügbaren Bereichen auf Energieeffizienz geachtet (inkl. Abwärmenutzung)?
- Welchen Einfluss auf das Klima haben diese Technologien?

Datenanalyse

Bei der Datenanalyse werden Informationen aus vorhandenen Daten gewonnen. Hier greift man oft auf anspruchsvolle Auswerterroutinen zurück. Hilfreich ist zum Beispiel die Nutzung von hybriden Modellen: die reduzierte Zahl der benötigten Simulationen hilft den Energiebedarf zu minimieren. In Zukunft können neue Ansätze wie Quantencomputer auch zu einer starken Reduktion von Rechenzeiten und somit Energieersparnis führen. Zusätzlich kann Förderung von Open-Source Projekten, also eine vereinfachte, gemeinsame Optimierung von Algorithmen, zu erhöhter Energieeffizienz führen.

Im Bereich Hardware für Datenanalysen soll im besten Fall auf energieeffiziente Systeme zurückgegriffen werden (z. B. LUMI Supercomputer in Kajaani, Finnland²⁵). Firmen wie Intel und Microsoft entwickeln spezielle, energieeffiziente Hardware für Privacy-preserving

¹⁹ [A critical analysis of Power Usage Effectiveness and its use in communicating data center energy consumption - ScienceDirect](#)

²⁰ [Usage impact on data center electricity needs: A system dynamic forecasting model - ScienceDirect](#)

²¹ [GreenDSS tool for data center management | IEEE Conference Publication | IEEE Xplore](#)

²² <https://www.ingenieur.de/fachmedien/hlh/raumluftechnik/kuehlung-von-rechenzentren/>

²³ https://www.ki-portal.de/wp-content/uploads/featured_image/37_wissen_roessel.pdf

²⁴ <https://heizung.de/heizung/wissen/sauberes-heizen-mit-serverabwaerme-vor-und-nachteile/>

²⁵ [United States Data Center Energy Usage Report | Energy Technologies Area \(lbl.gov\) https://www.lumi-supercomputer.eu/sustainable-future/](#)

Technologien. Virtualisierung oder auch die Nutzung von Cloud Ressourcen statt der Betreuung von nur einem Rechenzentrum stellen weitere energieschonende Alternativen dar.

Im Jahr 2020 sind weltweit schätzungsweise 18 Millionen Server in Rechenzentren im Einsatz gewesen.²⁶ Die Leistung der CPU-Sockets (Central Processing Unit) ist seit 2007 konstant geblieben und beträgt 118 W für Single-Socket-Server und 365 W für zwei Socket-Server.^{27, 28}

Die heute verfügbare Rechenleistung ist die treibende Kraft für effiziente (künstliche Intelligenz) Algorithmen. Computer Vision ist das Paradebeispiel für einen Bereich, der stark davon profitiert hat. Gerade in diesen Bereichen gibt es auch einen spezialisierten Bedarf an Prozessoren. Es werden mehr und mehr Graphics Processing Units (GPU) statt Central Processing Units (CPU) benötigt.

Dieses Beispiel soll vereinfacht die Energiekosten bei der Nutzung von GPU Ressourcen darstellen. Hierbei geht man von einem 1:1 Verhältnis von Grafikkarte zu Prozessor aus.

Wie hoch sind die Energiekosten beim Betrieb (unter Leerlauf/Durchschnitt/Volllast) einer gängigen Grafikkarte wie der Nvidia Tesla Serie? Auf der Nvidia-Seite heißt es, dass die maximale Leistung 250 W beträgt. Das ist etwa 2-mal größer als die Leistung einer aktuell durchschnittlichen Intel Xeon CPU (Xeon Gold 16-Core beispielsweise bis 165 W). Die Leistung pro Tag beträgt dann $24 \cdot 250 = 6$ kWh für GPU und $24 \cdot 165 = 3,9$ kWh für CPU.

Das bedeutet für ein Rechenzentrum mit 5k GPUs bei Volllastbetrieb:

Wie viel Energie kostet ein solches Rechenzentrum nach unseren bisherigen Zahlen pro Tag unter der Annahme, dass pro System 4 GPUs verbaut sind?

$5k \cdot 6$ kWh für GPU und $5k/4 \cdot 3.9$ kWh für CPU. In Summe also rund 35 MWh pro Tag. Diese Berechnung enthält noch keine Verlustleistung oder weitere wenngleich geringfügigere Energieaufnahmen, die durch andere Systemkomponenten erfolgt. Zusätzlich tragen neben der Hardware auch andere Aspekte wie Kühlung erheblich zum Energieverbrauch bei. Die tatsächliche Energieaufnahme kann jedoch je nach Betrieb stark variieren und muss system- oder raumspezifisch vermessen werden.

Relevante Fragestellungen

- Welche Speicher- und Rechenressourcen werden für die Analysen verwendet?

²⁶ [United States Data Center Energy Usage Report | Energy Technologies Area \(lbl.gov\)](#)

²⁷ <https://doi.org/10.1088/1748-9326/aaec9c>

²⁸ [United States Data Center Energy Usage Report | Energy Technologies Area \(lbl.gov\)](#)

- In welchen Fällen wird auf Virtualisierung und Cloud Computing zurückgegriffen?
- Mit welcher Energieform wird die Cloud angetrieben (Erneuerbare)?
- Welche Algorithmen werden für die Analysen verwendet?

Datenaufbewahrung

Nachdem die Daten analysiert wurden und somit Information generiert wurde, werden die Daten aufbewahrt. In vielen Fällen werden dazu interne Speicherressourcen verwendet, oder auch Cloud Ressourcen. Hier empfiehlt sich eine Berechnung, welche Form der Speicherung höhere Energieeffizienz aufweist.

Die Leistung der Festplatte (HDD) wurde 2006 auf 14 W/Festplatte geschätzt, was jedes Jahr um ca. 5 Prozent sinkt (8,6 W/Festplatte im Jahr 2015). Die Leistung der Solid State Disk (SSD) ist seit 2010 konstant bei 6 W/Festplatte geblieben, aber die Leistung pro Terabyte (TB) hat sich verbessert. Der Gesamtstromverbrauch für Festplatten in den USA im Jahr 2020 wird auf etwas mehr als 8.000 GWh/Jahr bei insgesamt 1.000 Millionen TB Speicher geschätzt. Mit einer geschätzten Lebensdauer von 4,4 Jahren ist die Anzahl der eingesetzten Festplatten stabil, aber die Gesamtkapazität steigt.²⁹

Emissionsberechnungen physischer Infrastruktur werden im Greenhouse Gas Protocol³⁰ beschrieben. Diese Arten von Emissionen fallen unter die Scope-1- und Scope-2-Berichtsrichtlinien, die von vielen Organisationen veröffentlicht werden müssen.³¹ Zusätzlich existieren Standards für den Bau von energieeffizienten Rechenzentren.³²

Wenn Cloud-Ressourcen genutzt werden, gibt man damit verbundenen Emissionen in der freiwilligen Scope-3-Berichterstattung als „indirekte“ oder ausgelagerte Emissionen an. Berechnung der tatsächlichen Emissionen gestaltet sich schwieriger, da Cloud-Anbieter*innen wie Amazon Web Services, Google Cloud und Microsoft Azure hauptsächlich aggregierte globale Daten mit unterschiedlicher Transparenz veröffentlichen.^{33,34} Dieser Mangel an Transparenz führte auch dazu, dass Amazon in einem Greenpeace-Bericht kritisiert wurde.³⁵

Parallel dazu setzen sich die skalierbaren Infrastrukturen auch für Energieeffizienz ein, beispielsweise durch den Bau eigener Server durch Google^{36,37} und den Bau des ersten

²⁹ [United States Data Center Energy Usage Report | Energy Technologies Area \(lbl.gov\)](#)

³⁰ [ghg-protocol-revised.pdf \(ghgprotocol.org\)](#)

³¹ [Streamlined energy and carbon reporting - GOV.UK \(www.gov.uk\)](#)

³² [Energy Efficient Data Centers | SpringerLink](#)

³³ [Unsere CO2-Bilanz \(aboutamazon.de\)](#)

³⁴ [Assessing the suitability of the Greenhouse Gas Protocol for calculation of emissions from public cloud computing workloads | Journal of Cloud Computing | Full Text \(springeropen.com\)](#)

³⁵ [Greenpeace Report: Click Clean Virginia - Greenpeace USA](#)

³⁶ [Übersicht über das Sicherheitsdesign der Infrastruktur von Google | Lösungen | Google Cloud](#)

³⁷ [Google's Making Its Own Chips Now. Time for Intel to Freak Out | WIRED](#)

Gasrechenzentrums durch Microsoft.³⁸ Der Technologiesektor ist auch der größte Abnehmer von erneuerbaren Energien.^{39,40}

In der Aufbewahrungsphase des Datenlebenszyklus werden Daten auch mit Kooperationspartner*innen ausgetauscht. Zusätzliche Datenströme durch gesteigerten Datenaustausch bewirken allerdings auch einen erhöhten Energieverbrauch. Daher ist es wichtig durch strategische Ansätze den Datenaustausch nachhaltig zu gestalten (z.B. optimierter Ressourceneinsatz durch Föderation). Unternehmen sollten in diesem Bereich auch auf effiziente Datenverwaltung setzen (z.B. Informationslebenszyklusmanagement), unnötige Cloud-Daten löschen und Datenprofile nutzen.

Relevante Fragestellungen

- Welche Speicherressourcen werden für die Aufbewahrung der Daten verwendet?
- In welchen Fällen wird auf Virtualisierung und Cloud-Speicherressourcen zurückgegriffen?
- In welchen Fällen wird auf eine föderierte Infrastruktur zurückgegriffen?

Daten (Nach-)Nutzung und Datenvernichtung

In der letzten Phase des Datenlebenszyklus endet jedoch nicht die Nutzbarkeit der Daten. Im Optimalfall werden die Daten für neue Fragestellungen nachhaltig genutzt. In einigen Unternehmen gibt es jedoch große Datenmengen, die auf Speicherressourcen verbleiben und nicht mehr genutzt werden. In vielen Fällen ist das die günstigere Variante, da die Bewertung der Nutzbarkeit der Daten ein neuer Prozess ist, der spezielle Humane Ressourcen erfordert.

Um die gemeinsame Nutzung von Daten zu maximieren, wurde eine Vielzahl von Initiativen ins Leben gerufen. Auf europäischer Ebene spielen die International Data Spaces⁴¹ eine wichtige Rolle. Hier werden Datenökosysteme in unterschiedlichen Domänen etabliert, um die gemeinsame Datennutzung voranzutreiben. Auf nationaler Ebene hat die Data Intelligence Offensive⁴² gemeinsam mit dem Bundesministerium für Klimaschutz, Umwelt, Energie, Mobilität, Innovation und Technologie eine domänenspezifische Initiative zur Nutzung von Daten im Bereich Energiewende, Mobilitätswende, Kreislaufwirtschaft und Klimawandel Mitigation, den Green Data Hub⁴³, ins Leben gerufen. Zusätzlich gibt es auch regionale Ansätze,

³⁸ [Redesigning Datacenters for an Advanced Energy Future - Microsoft Green Blog](#)

³⁹ [Data centres and energy – from global headlines to local headaches? – Analysis - IEA](#)

⁴⁰ [Microsoft will be carbon negative by 2030 - The Official Microsoft Blog](#)

⁴¹ [International Data Spaces | The future of the data economy is here](#)

⁴² [Start - DIO – Data Intelligence Offensive](#)

⁴³ [Start - Green Data Hub](#)

um die kollaborative Datennutzung voranzutreiben, wie das steirische Projekt Innovative Data Environment @ Styria⁴⁴.

Relevante Fragestellungen

- Welche Sichtbarkeit haben die Daten, um mögliche Nachnutzung zu maximieren?
- Welche Prozesse gibt es zur effizienten Daten (Nach-)Nutzung?
- Welche Prozesse zur Vernichtung von Daten werden angewandt?

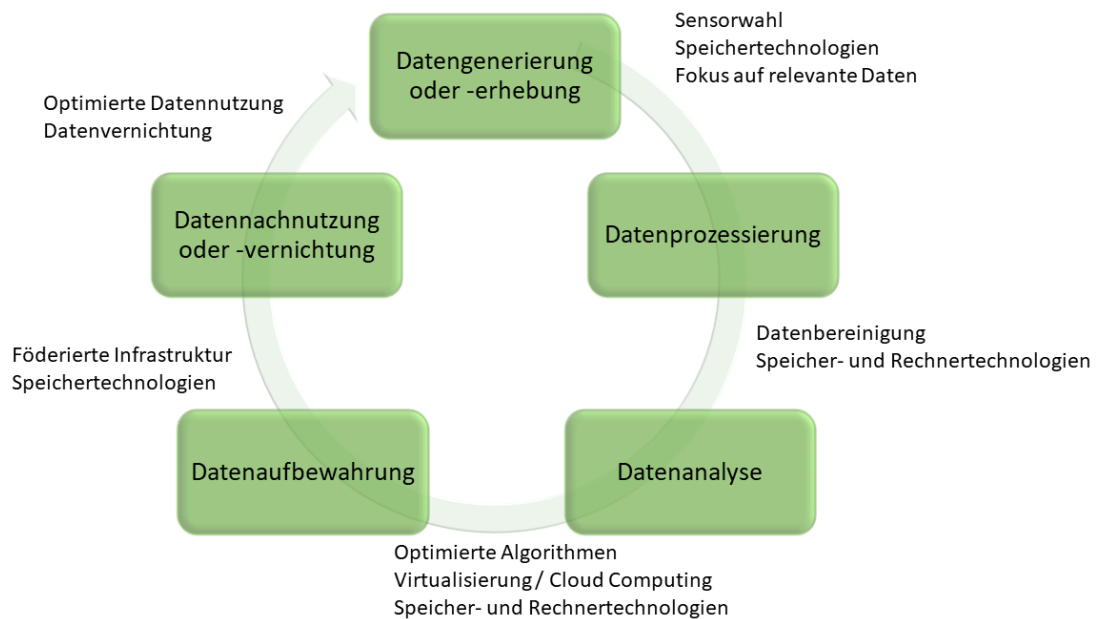


Abbildung 1: Datenlebenszyklus und Nachhaltigkeitsaspekte

Weitere Nachhaltigkeitsaspekte für Daten

Zusätzlich zu den Aspekten um den Datenlebenszyklus können weitere Ansätze zur Entwicklung zu Green Data helfen.

Die Nutzung von energieeffizientem Equipment (z.B. Serverdimensionierung, Blade Server, Passiv statt Aktivkühler, Thermal Design Power CPUs, OLED Displays, SSD statt HDD), langjährige Nutzungsdauer, ökologische Bilanz in der Herstellung sowie das Recycling von IT Equipment ist ein weiterer umweltschonender Ansatz.

Es werden jährlich Milliarden von elektronischen Geräten produziert, verkauft und wieder entsorgt. Das bringt einen großen Profit für die Elektronikhersteller*innen, geht jedoch auch

⁴⁴ [IDEAS - Home \(tugraz.at\)](http://IDEAS - Home (tugraz.at))

einher mit einer großen Belastung für unseren Planeten. Hier sollte vor allem an die Nutzung von erneuerbaren Rohstoffen gesetzt werden. Wichtig sind außerdem Herkunft der Rohstoffe – seltene Erden, Lithium, Gold, Kupfer usw. – sowie der Verzicht auf den Einsatz umweltgefährlicher Chemikalien wie PVC oder bromierte Kohlenwasserstoffe. Außerdem hat die geplante Obsoleszenz von Produkten einen großen Einfluss auf die Umwelt.⁴⁵

Im Bereich Humanressourcen kann die Reduktion von Dienstreisen durch effiziente online Kollaborationsmöglichkeiten auf Datenebene einen Beitrag zur Ressourcenschonung leisten. Zusätzlich ermöglichen mobile Arbeitsplätze sowie z.B. auch solarbetriebene Kleinanlagen am Rucksack und Schulungen zum Thema Energieeffizienz Abhilfe und erhöhen das Bewusstsein für grüne Daten.

Zusätzlich können die Daten auch für einen nachhaltigen Anwendungszweck erhoben werden (z. B. datengetriebene Optimierung von Abfallverwertungsprozessen, Datengetriebene Optimierung von Energienutzung).

Zusammenfassung

Diese Publikation beschreibt jene Prozesse im Lebenszyklus von Daten, die für Nachhaltigkeit und Grüne Daten eine zentrale Rolle spielen. Ein großes Augenmerk wurde hierbei auf Energieeffizienz gelegt: Welche Energieform wird genutzt? Wird auf nachhaltige Energieformen gesetzt? Wird die Abwärme bei Energiegewinnung und -einsatz genutzt? Werden energieeffiziente Technologien verwendet?⁴⁶

Im nächsten Schritt müssen für diese Aspekte Metriken definiert werden, um eine Messbarkeit zu ermöglichen. Diese Metriken können dann eine Messgrundlage für eine Standardisierung bieten. Dies ermöglicht in Zukunft ein Gütesiegel für grüne Daten auszustellen. Gütesiegel helfen Verbraucher*innen bei der Bewertung von Produkten. Produkte aus ökologischem Anbau werden mit einem BIO-Siegel gekennzeichnet, Lebensmittel, die unter bestimmten sozialen und ökologischen Kriterien produziert wurden, werden mit dem Fairtrade Siegel markiert. Ein solches Gütesiegel wäre auch im Bereich Datenwirtschaft denkbar. Mit dem Selbstbewertungsfragebogen im Annex (Tabelle 1) möchten wir dazu beitragen dieses Ziel in Zukunft erreichen zu können. Dieser Fragebogen behandelt jene Aspekte, die für grüne Daten relevant sind, und ermöglicht Unternehmen und Institutionen sich mit den Aspekten im Detail auseinanderzusetzen. Das ist ein erster Schritt in die Richtung grüne Datenwirtschaft in Österreich.

⁴⁵ <https://www.greenpeace.de/sites/default/files/publications/20171016-greenpeace-guide-greener-electronics-englisch.pdf>

⁴⁶ https://www.greenpeace.de/sites/default/files/publications/20170110_greenpeace_clicking_clean.pdf

Um diese Prozesse auch nachhaltig in Unternehmen zu integrieren, sollten diese Metriken in unternehmensinternen Systemen integriert werden und mit Recommender Systemen Empfehlungen für die Ökologisierung der Daten gegeben werden.

Danksagung

Wir danken recht herzlich Nunu Kaller und Dr. Hermann Schranzhofer für wertvolle Anregungen und Hinweise für die Verfassung dieser Publikation.

Autor*innen



Dr. Sarah Stryeck – Senior Researcher (Know-Center GmbH), PostDoc (Technische Universität Graz), Data Steward (DIO). Sarah koordiniert das Projekt Innovative Data Environments @ Styria, leitet im Projekt EuroCC die Task Competence Mapping für HPC, Big Data und AI und befasst sich mit den Themen Datenplattformen und Vertrauenswürdige KI.

Dr. Claire Jean-Quartier – PostDoc & Data Steward (Technische Universität Graz). Claire ist Forschungsassistentin an der TU Graz und bringt Erfahrung aus Qualitätsmanagement, Bioinformatik und Grundlagenforschung mit, die alle Schritte des Datenzyklus einbezieht.



DI Mag. Günther Tschabuschnig - Präsident (DIO) – Günther leitet die Data Intelligence Offensive und legt seinen Fokus auf Data Intelligence, Data Driven Buisness development und nachhaltige Datenverarbeitung.

Kontakt für Rückfragen:

DIO – Data Intelligence Offensive

Thurngasse 8, 1090 Wien

Sarah Stryeck

Sarah.Stryeck@dataintelligence.at

Annex: Selbstbewertungsfragebogen

Phase im Datenlebenszyklus	Fragestellungen	Relevante Aspekte	Wie kann man entgegenwirken?
Datengenerierung/-erhebung	Wie werden die Daten erzeugt?	Daten werden durch Sensoren generiert <ul style="list-style-type: none"> • nachhaltige Produktion oder recyceltes Material • black Plastic • Lebensdauer > 1 Jahr • energieeffizient? • Grüner Strom für Betreibung • brauner Strom für Betreibung 	Infrastrukturwahl
		Manuelle Erhebung	Prozessoptimierung
	Werden Daten wiederverwendet?	Daten werden aus einer API bezogen	Infrastrukturwahl
		Einmalige Messung	Prozessoptimierung
Datenpreprozessierung	Welche Prozesse gibt es zur Entfernung von unnötigen Datenmengen?	Personal zur Bewertung von Datensätzen	Prozessoptimierung
	Welche Speicher- und Rechnertechnologien werden für die Verarbeitung verwendet?	Grüner Speicher oder grüner Strom für Betrieb. CO ₂ -neutrales Rechenzentrum	Infrastrukturwahl
	Haben die eingesetzten Technologien einen negativen Einfluss auf das Klima?	Grüner Strom für eingesetzte Technologien	Infrastrukturwahl

Datenanalyse	Welche Rechenressourcen werden für die Analysen verwendet?	lokale Ressourcen (SSD, HDD) Virtualisierung / Cloud Computing	Infrastrukturwahl
	Welche Algorithmen werden für die Analysen verwendet?	optimierte Algorithmen (Energieeffizienz)	Prozessoptimierung
	Im Falle der Nutzung von Data Centers: Welches Kühlungssystem wird verwendet?	Serverraum-Klimaanlagen, Immersionskühlung, Kaltgang-/Warmgangssystem, freie Kühlung, Adiabatische Kühlung,..	Infrastrukturwahl
Datenaufbewahrung	Welche Speicherressourcen werden für die Aufbewahrung der Daten verwendet?	lokale Ressourcen (SSD, HDD, Band, Glacier Storage) Virtualisierung / Cloud Computing	Infrastrukturwahl
	Wird auf optimierte Ressourcennutzung gesetzt?	föderierte Ressourcen	Infrastrukturwahl
Daten (Nach-)Nutzung und Datenvernichtung	Welche Sichtbarkeit haben die Daten, um mögliche Nachnutzung zu maximieren?	Öffentliche Repositorien / Open Data Hubs Social Media Bewerbung	Prozessoptimierung
	Welche Prozesse gibt es zur effizienten Daten (Nach-)nutzung?	Schnittstellen zu öffentlichen Repositorien / Open Data Hubs Interne Prozesse zur Nutzung von Public Data	Prozessoptimierung
	Welche Prozesse zur Vernichtung von Daten werden angewandt?	Datenbewertung auf Nutzbarkeit / Vollständigkeit / Qualität / Plausibilität	Prozessoptimierung