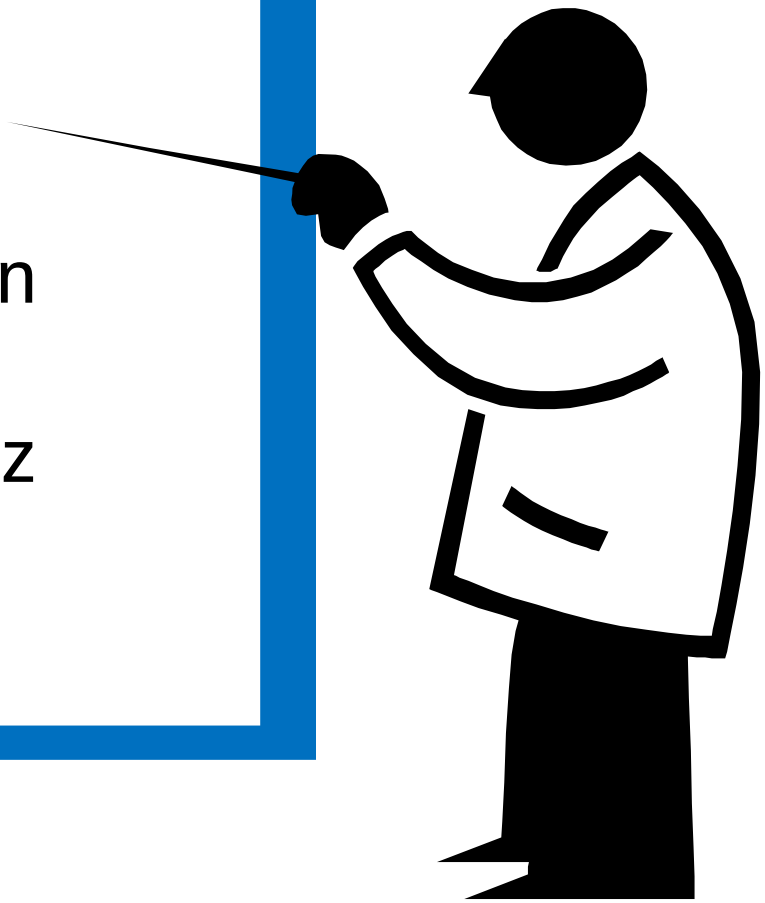
Sicherheits- und Datenschutzaspekte von Big Data

Herbert.Leitold@a-sit.at

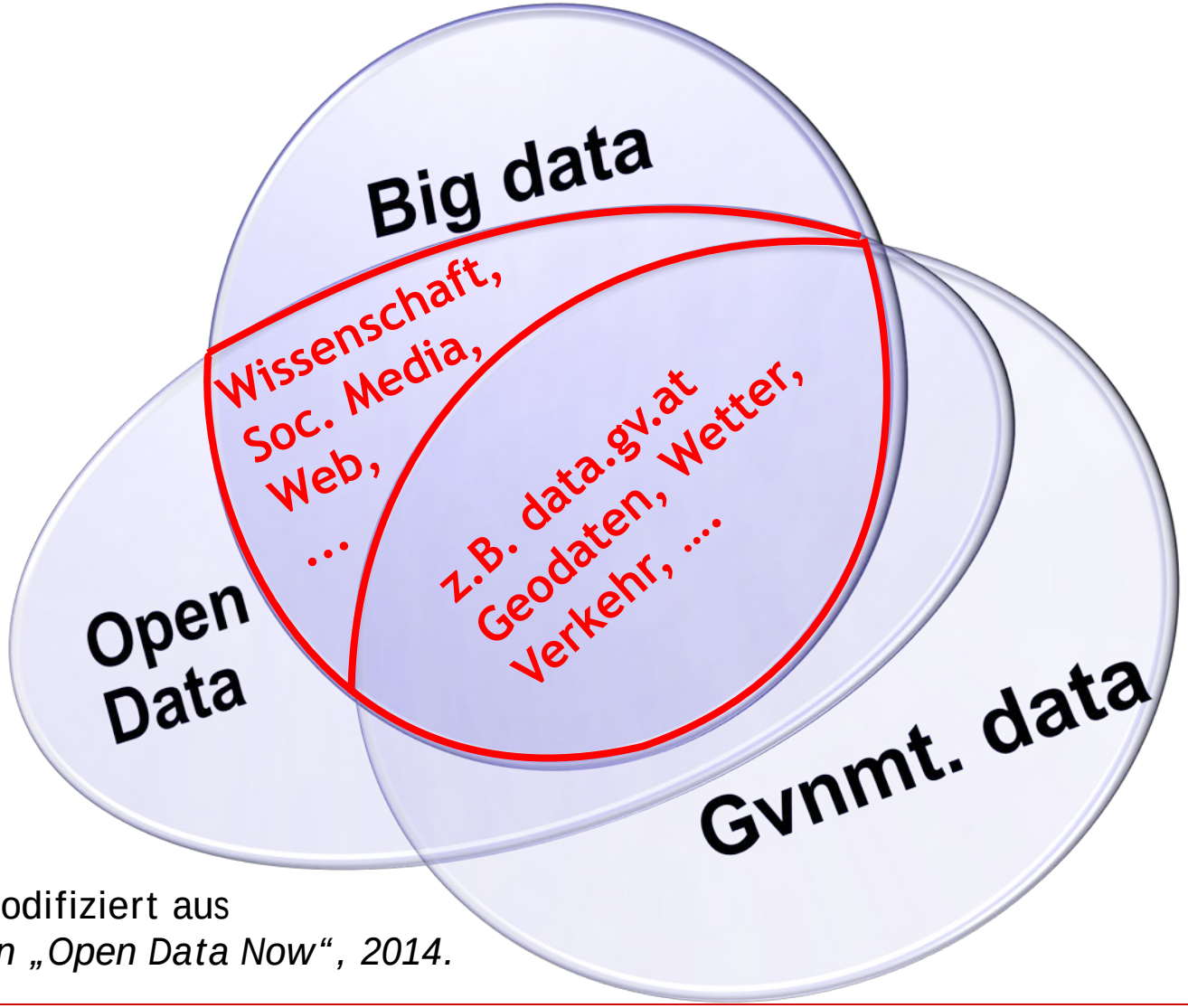
ADV-Tagung Big Data 2016

Übersicht

- Einleitung
- Sicherheitsanforderungen
- Big Data und Datenschutz
- Forschungsprojekte



Umfeld und Fokus des Vortrags

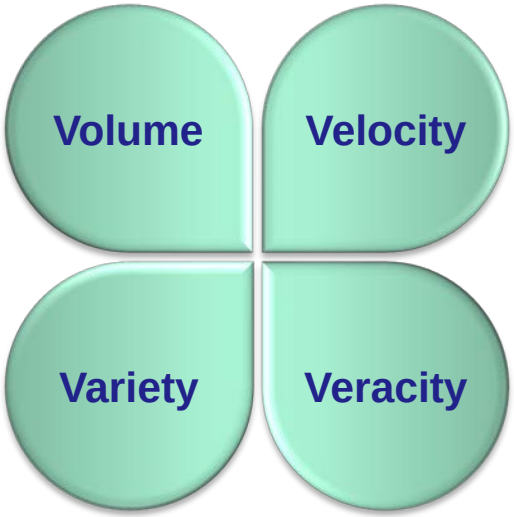


Quelle: Modifiziert aus
Joel Gurin „Open Data Now“, 2014.

Ist „Big Data“ speziell bzgl. Sicherheit?

- Ergeben sich aus Big Data Charakteristika spezifische Sicherheits-Anforderungen?

- Big Data Charakteristika
  - *Volume* – Datenmenge
  - *Velocity* – kurze Lebensdauer
  - *Variety* – Quellen, Datenarten
  - *Veracity* – Richtigkeit



Quellen


cloud security alliance
Top Ten Big Data Security and Privacy Challenges
CSA, 2012
e-GOVERNMENT BUND-LÄNDER-GEMEINDEN
http://reference.e-government.gv.at
White Paper
Big Data – 1.0.0
Ergebnis der PG
Positionspapier zu Big Data in der öffentlichen Verwaltung
(Stand 06.06.2016)
Kurzbeschreibung
Der exponentielle
Maschine erzeugten
öffentliche Verwaltu
Chancen mit sich.
BLSG, 2016
ARTICLE 29 DATA PROTECTION WORKING PARTY
14/EN
WP 221
Statement on Statement of the WP29 on the impact of the development of big data on the protection of individuals with regard to the processing of their personal data in the EU
Art. 29, 2014

Übersicht

- Einleitung
- **Sicherheitsanforderungen**
- Big Data und Datenschutz
- Forschungsprojekte

Top Ten nach CSA

1. Secure Computations in Distributed Programming Frameworks
2. Security Best Practices for Non-Relational Data Stores
3. Secure Data Strorage & Transaction Logs
4. End-Point Input Validation/Filtering
5. Real-time Security/Compliance Monitoring
6. Scalable & Composable Privacy-Preserving Data Mining & Analytics
7. Cryptographically Enforced Access Control & Secure Communication
8. Granular Access Control
9. Granular Audits
10. Data Provenance

Secure Computation / NoSQL

1. Rechnen in verteilter Umgebung
   - Massive Parallelisierg. zB *MapReduce Framew*.
   - Kompromittierter Mapper kann falsche Daten liefern, daraus falsche aggregierte Ergebnisse
   - Mapper könnten vertrauliche Daten leaken
2. NoSQL Security Best Practices
   - Entwicklung aus Analyse-Sicht, Sicherheit meist kein Schwerpunkt der Entwicklung

Secure Storage / Input Validation

3. Sichere Speicherung und Transaction Logs
   - Multi-tiered storage benötigt auto-tiering für storage management. Wo liegen die Daten?
   - Rechtsraum-bewusstes Speichern (legislation-aware)?
4. End-point Input-Validation, Filtering
   - Hohe Zahl an Sensoren, BYOD, …
   - Authentizität der Quelle, zB rogue virtual sensors

Real-Time Monitoring / PET

5. Echtzeit Security/Compliance Monitoring
   - Über hohe Volumen herausfordernder
   - Mehr Daten und Analyse kann schnellere und bessere Einschätzung/Entscheidung bringen
6. Skalierbare Datenschutz-wahrende Daten-Sammlung und -Analyse
   - Aus besserer Korrelation potentiell stärkerer Personenbezug und Profile

Kryptographie / Granularer Zugang

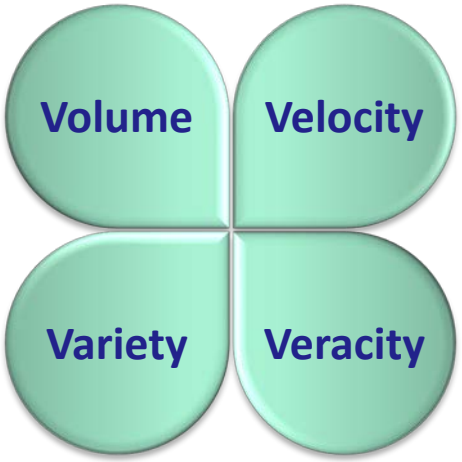
7. Kryptographie-gestützte Zugangskontrolle
   - Traditionelle Verschlüsselung „alles oder nichts“
   - Forschung zu Attribute-Based Encryption
   - Rechnen mit Secure Multiparty Communication
8. Granulare Zugangskontrolle
   - Traditionell grob-granular; höhere Granularität macht Management aufwändiger
   - Automatisierung policy-enforcement

Granulares Audit / Provenance

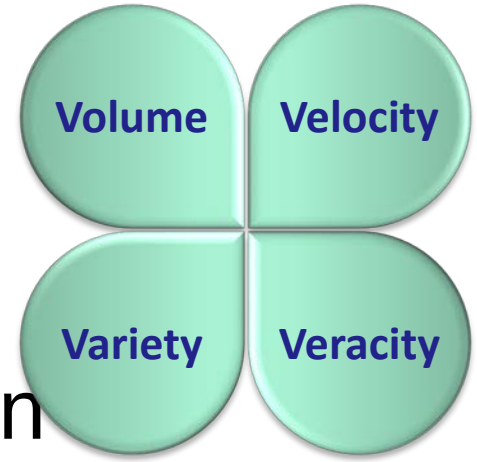
9. Granulare Audits
   - Für Audit/Forensik mehr Datenquellen, die potentiell verteilt sind
   - Audit-Daten aus Compliance-Anforderungen
10. Datenherkunft
   - Metadaten zur Herkunft steigen im Volumen
   - Anforderungen aus Compliance, zB Insider-Handel

Aus Big Data Eigenschaften

- Wie stehen die diskutierten Herausforderungen zu Big Data Charakteristika?
- Was macht Big Data spezifisch?

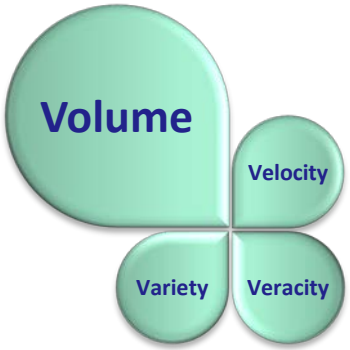


Herausforderung aus Volumen



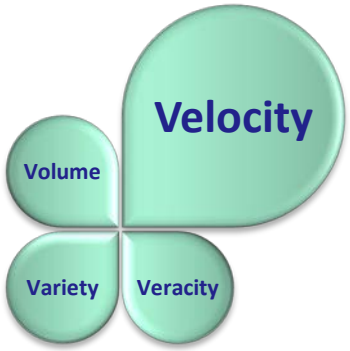
- Sicheres Rechnen in verteilten Systemen
- Rechtsraum-bewusstes Speichern
- Anforderungen an Effizienz von krypto-graphischen Verfahren, aber auch an die Funktionalität (z.B. ABE, SMPC)
- Granulare Audits oder Zugangskontrolle

Herausforderung aus Lebensdauer



- Echtzeit Security-Montoring
- Nachweis zu Datenherkunft
- Anforderungen an Effizienz von krypto-graphischen Verfahren

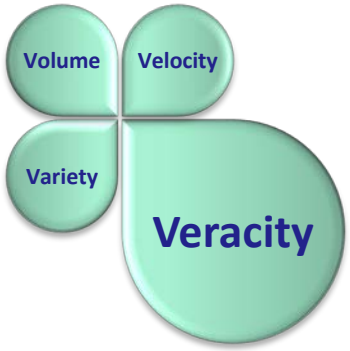
Herausforderung aus Richtigkeit



- Authentizität der Daten von Mappern aus massiver Parallelisierung
- Authentizität der Datenquellen, z.B. von Sensoren oder BYOD

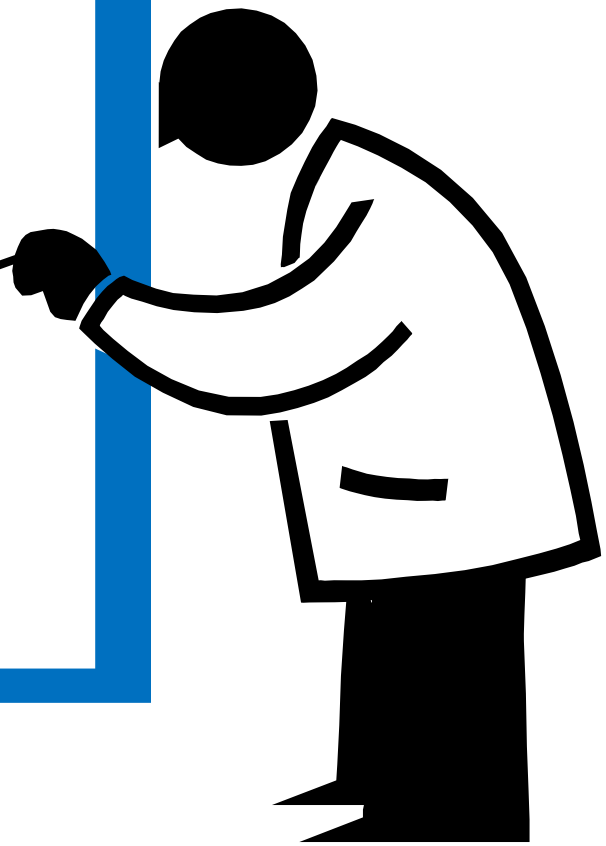
Herausforderung aus Vielfalt/Datenarten

- Input-validation bei hoher Anzahl (verschiedener) Quellen
- Mehr Daten und bessere Analyse zielt auf bessere Ergebnisse ab
- Damit aber auch potentiell höhere Qualität im Personenbezug => *Datenschutz*



Überblick

- Einleitung
- Sicherheitsanforderungen
- **Big Data und Datenschutz**
- Forschungsprojekte



Automatisierte Entscheidungen

- Big Data verspricht bessere (automatisierte) Entscheidungen

EU DSGV §22: Die betroffene Person hat das Recht, nicht einer ausschließlich auf einer automatisierten Verarbeitung — einschließlich Profiling — beruhenden Entscheidung unterworfen zu werden, die ihr gegenüber rechtliche Wirkung entfaltet oder sie in ähnlicher Weise erheblich beeinträchtigt.

Zweckbindung der Datenverarbeitung

- Bei Verknüpfung von Daten mit Personenbezug ist Zweckbindung ggf. verletzt

> **EU DSGV §5(b)**: Personenbezogene Daten müssen … für festgelegte, eindeutige und legitime Zwecke erhoben werden und dürfen nicht in einer mit diesen Zwecken nicht zu vereinbarenden Weise weiterverarbeitet werden

- Art. 29 WP sieht in Zweckbindung u.a. auch ein Vermeiden von Marktverzerrungen.

Sicherheit der Datenverarbeitung

- Verantwortliche der Datenquellen haben geeignete Schutzniveaus sicherzustellen

> **EU DSGV §21**: Unter Berücksichtigung des Stands der Technik, der Implementierungskosten und der Art, des Umfangs, der Umstände und der Zwecke der Verarbeitung sowie *[etc. etc.]* treffen der Verantwortliche und der Auftragsverarbeiter geeignete technische und organisatorische Maßnahmen …

- Bei mehreren Quellen *Maximum* einzelner Schutzniveaus *Minimum* der Verknüpfung

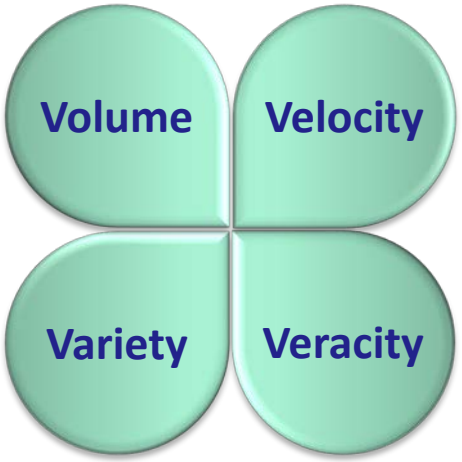
Anoymisierung, Pseudonymisierung

- Open Data oft ohne Personenbezug
  - Pseudonymisiert / Anonymisiert?

EU DSGV (26): Bei der Feststellung, ob Mittel nach allgemeinem Ermessen wahrscheinlich zur Identifizierung der natürlichen Person genutzt werden, sollten alle objektiven Faktoren, wie die Kosten der Identifizierung und der dafür erforderliche Zeitaufwand, herangezogen werden, wobei die zum Zeitpunkt der Verarbeitung verfügbare Technologie und technologische Entwicklungen zu berücksichtigen sind.

- Aus Vielzahl an Quellen ist möglich, dass urspr. nicht gegebener Personenbezug aus Verknüpfung hergestellt wird.

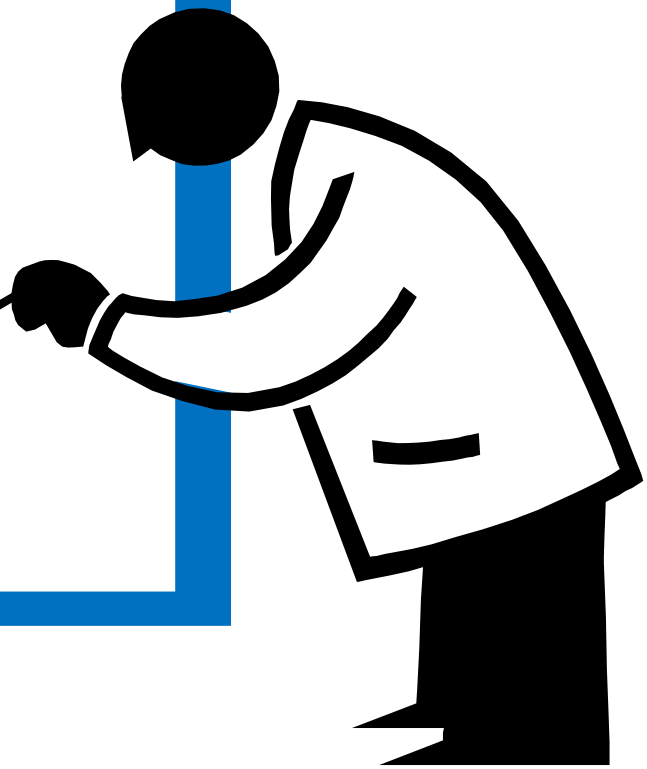
Zusammenfassung Datenschutz



- Vor allen Zusammenführung von Daten mehrerer Quellen ist zu bedenken
  - neuer Zweck der Datenverarbeitung
  - Personenbezug über Korrelation möglich

Überblick

- Einleitung
- Sicherheitsanforderungen
- Big Data und Datenschutz
- **Forschungsprojekte**



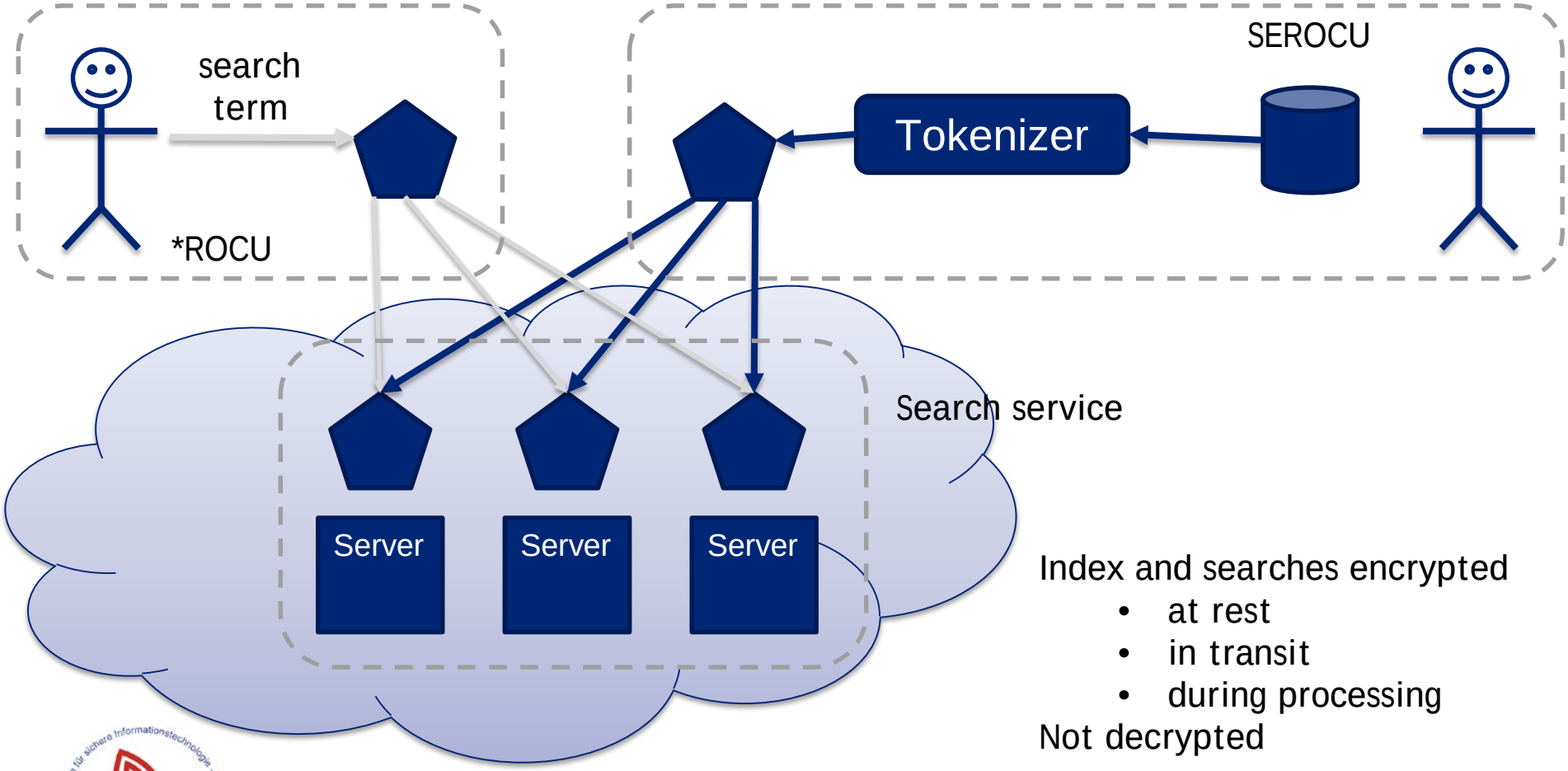
SUNFISH

- Horizon 2020 Forschungsprojekt zu föder-ieren von private Clouds der öff. Verwaltung
- Pilotierung in drei Use Cases
  - *Private Cloud Federation, with a Segregated Environment;*
  - *Hybrid Cloud Federation (Private and Public Cloud) a Segregated Environment;*
  - *Private Cloud Federation, with a Secure Multiparty Computation Environment.*

Use-Case „ROCUs“

- Regional Organised Crime Units (ROCUs) in UK halten große Mengen sensibler Daten
  - zB aus Beweissicherungen
- Lokale Such-Indizes in jeder ROCU, jedoch auch inter-ROCU Suche erforderlich, ohne Information über Quelle und Inhalt zu geben
- Unter Leitung von Cybernetica (EE) Such-Indizes für Secure Multiparty Communication gebildet
  - SMPC: Operiert auf verschlüsselten Daten

Such-Index mit SMPC



Index and searches encrypted

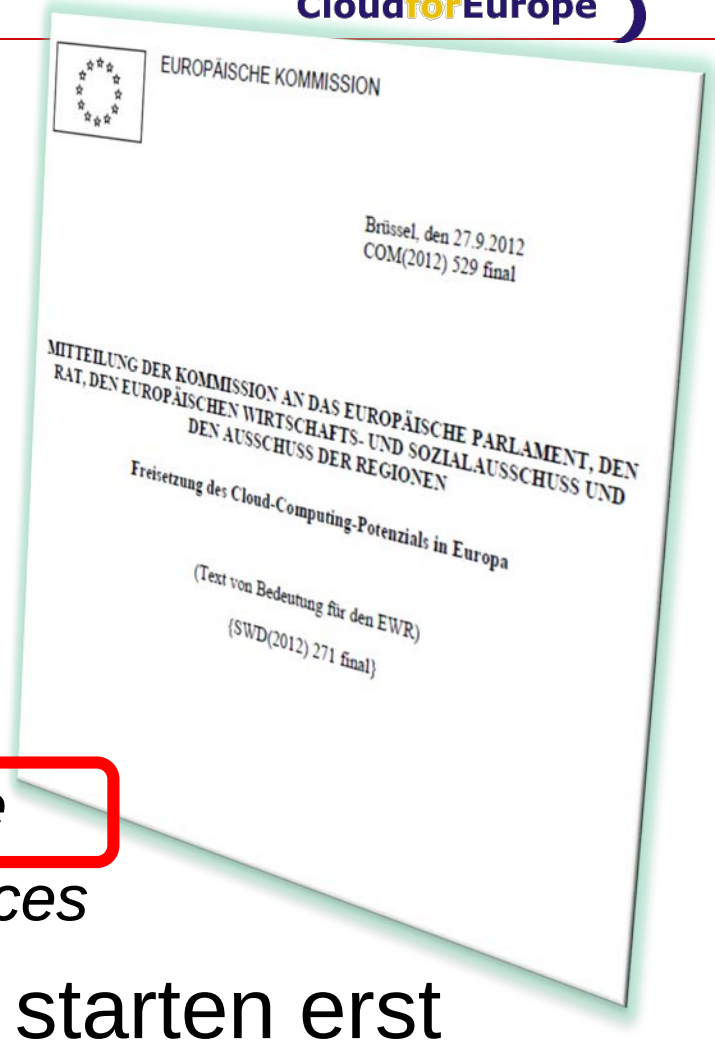
- at rest
- in transit
- during processing

Not decrypted

Cloud for Europe

- Vorkommerzielle Auftrags-vergabe zur Unterstützung der EK Mitteilung Cloud
- 3 Lose vergeben
  - *Federated Certified Services Brokerage for EU Public Administration Cloud*
  - *Secure Legislation-Aware Storage*
  - *Legislation Execution Cloud Services*

Arbeiten der Auftragnehmer starten erst


EUROPÄISCHE KOMMISSION

Brüssel, den 27.9.2012
COM(2012) 529 final

MITTEILUNG DER KOMMISSION AN DAS EUROPÄISCHE PARLAMENT, DEN RAT, DEN EUROPÄISCHEN WIRTSCHAFTS- UND SOZIALAUSSCHUSS UND DEN AUSSCHUSS DER REGIONEN

Freisetzung des Cloud-Computing-Potenzials in Europa

(Text von Bedeutung für den EWR)

{SWD(2012) 271 final}

Danke für Ihre Aufmerksamkeit!

Herbert.Leitold@a-sit.at

ADV-Tagung Big Data 2016