

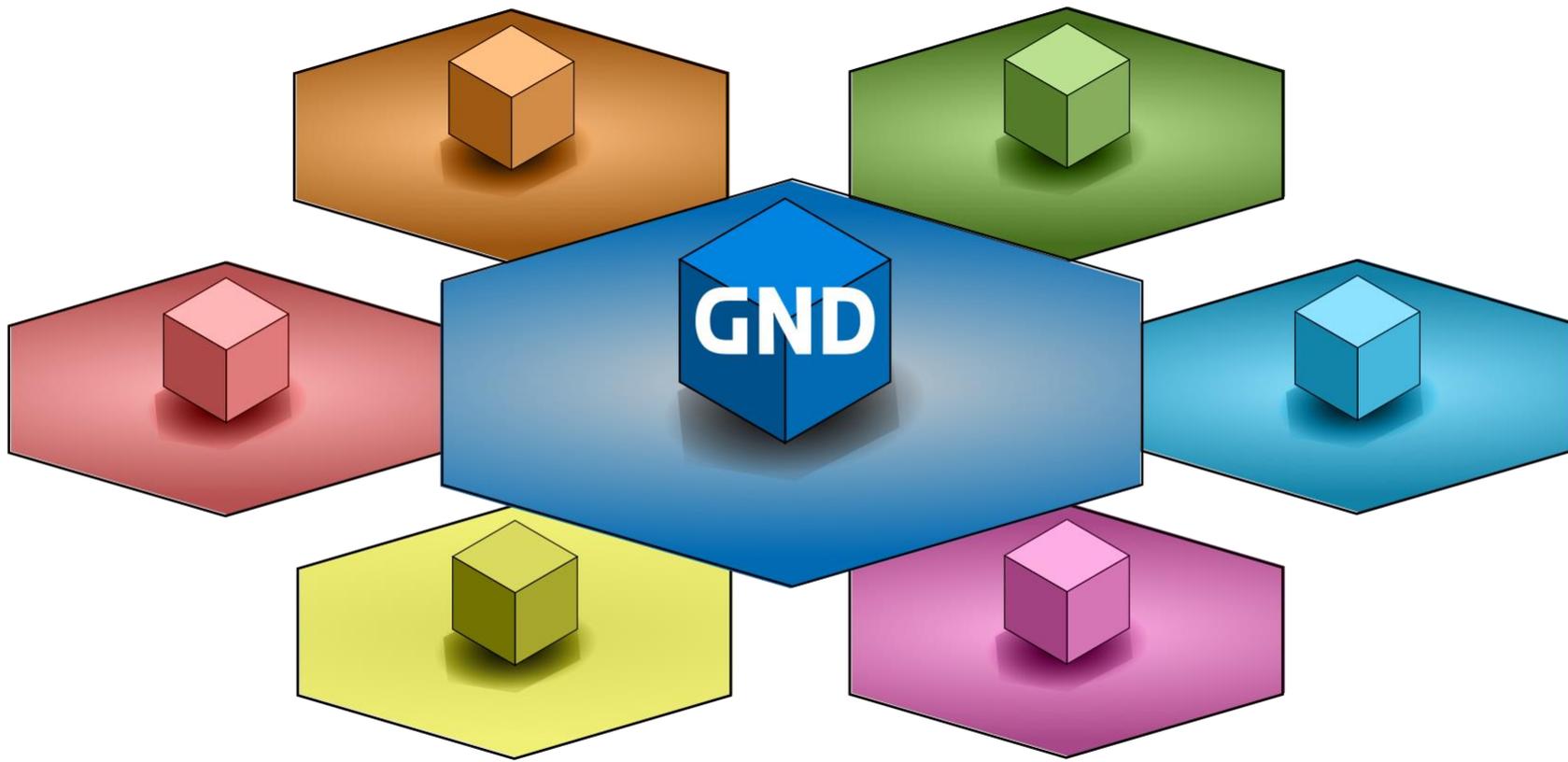
Jürgen Kett

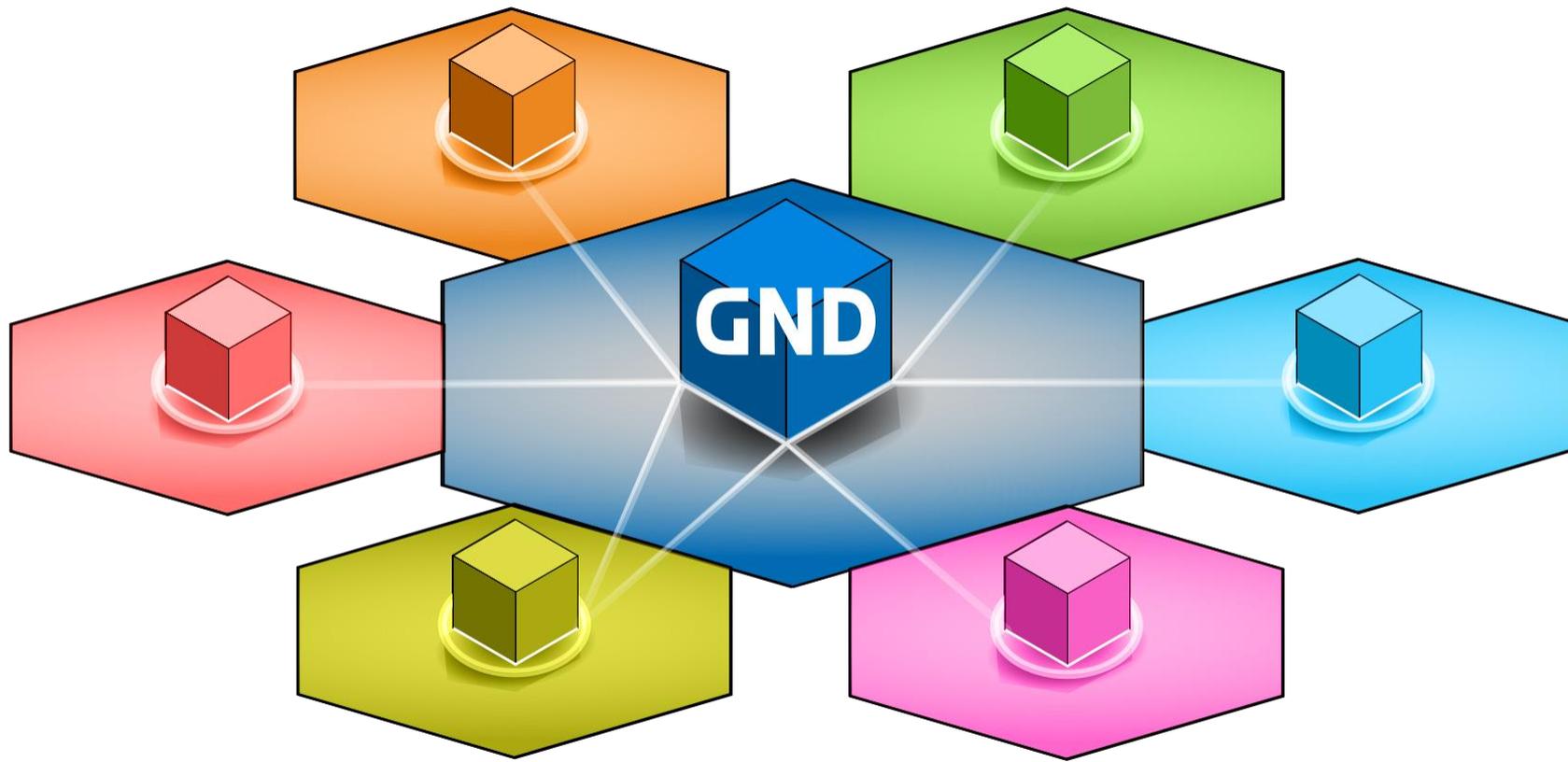
Datenqualität in der Gemeinsamen Normdatei (GND)

Kein Platz für Katalogisierungsautomaten?

Aufbau

- 1. Zur GND**
- 2. Problemstellung und Fragen**
- 3. Thesen und Diskussion**



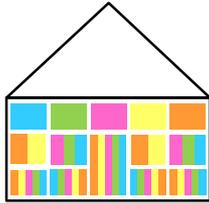




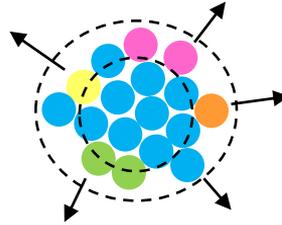
BRÜCKEN IM NETZWERK DER
KULTUR UND WISSENSCHAFT



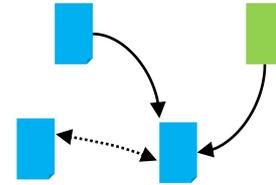
Organisation &
Kommunikation



Datenbasis



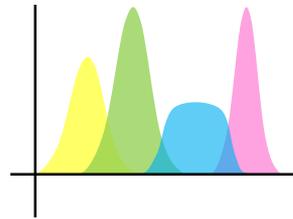
Vernetzung



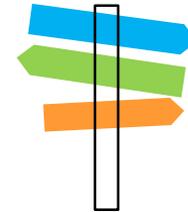
Regeln, Format &
Datenverwaltung



Redaktion &
Kooperation



Nutzung



Weitere Informationen unter: <https://gnd.network>

Was zeichnet die GND aus?



Verlässlichkeit

Wie versucht man Verlässlichkeit in der GND herzustellen?



ORGANISATION



INFRASTRUKTUR



INFRASTRUKTUR



STANDARDS
und REGELN



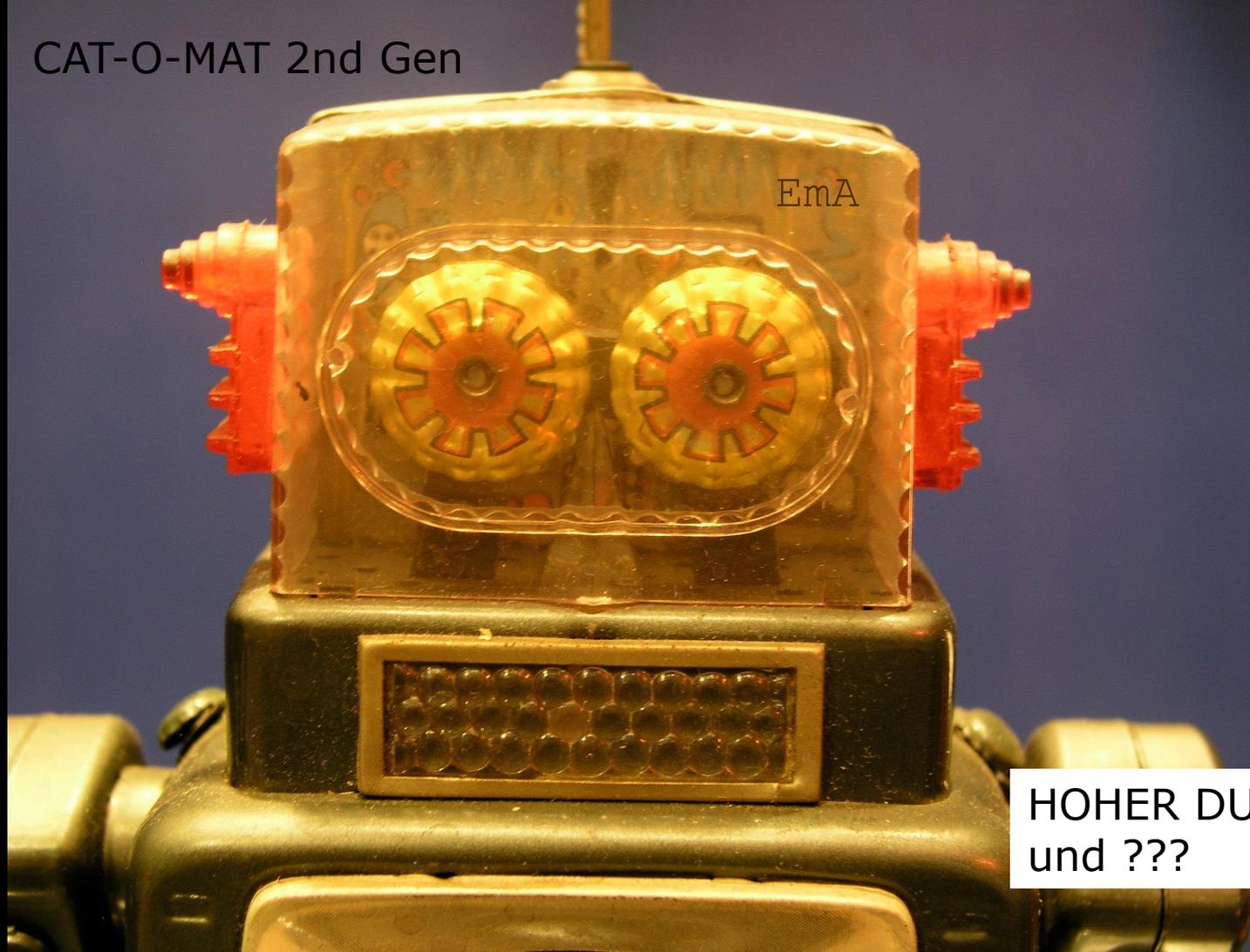
EXPERTISE und
QUALITÄTS-
BEWUSSTSEIN

*CJS*64*



CJS'64

CAT-O-MAT 2nd Gen



EmA

HOHER DURCHSATZ
und ???



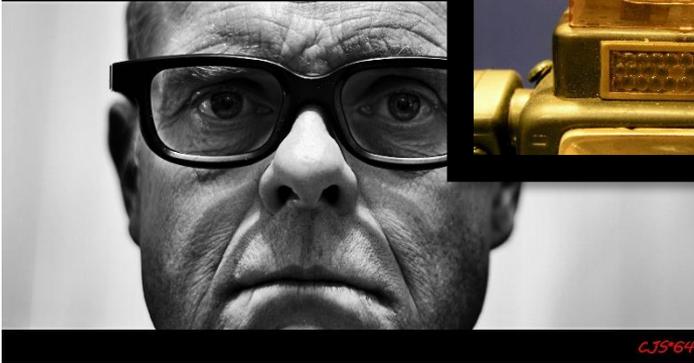
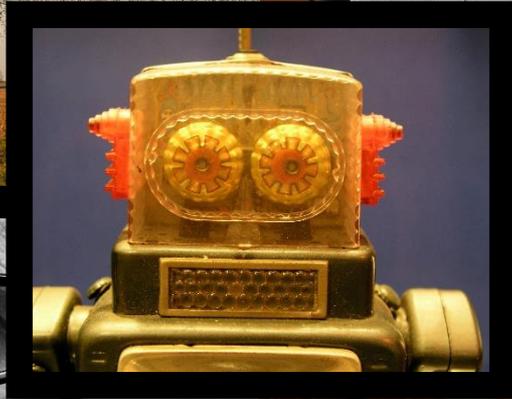
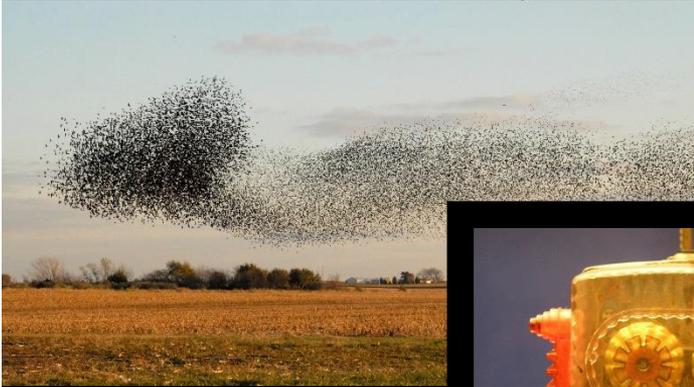
EXPERTISE und
QUALITÄTS-
BEWUSSTSEIN

*CJS*64*



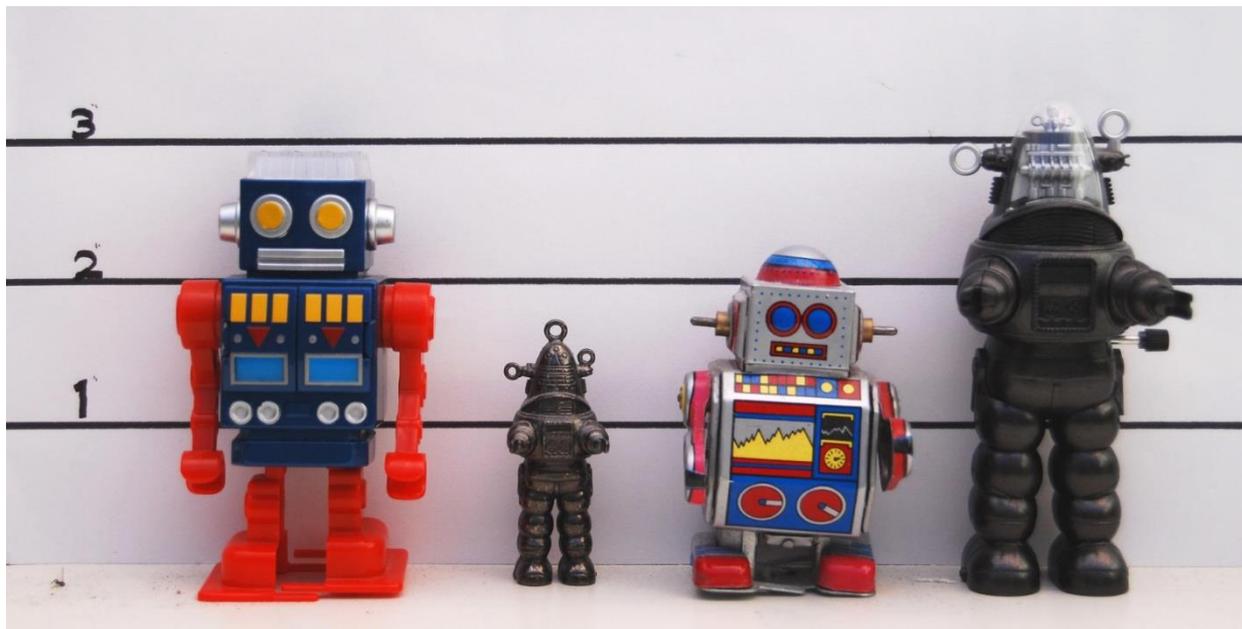
EmA

HOHER DURCHSATZ
und ???



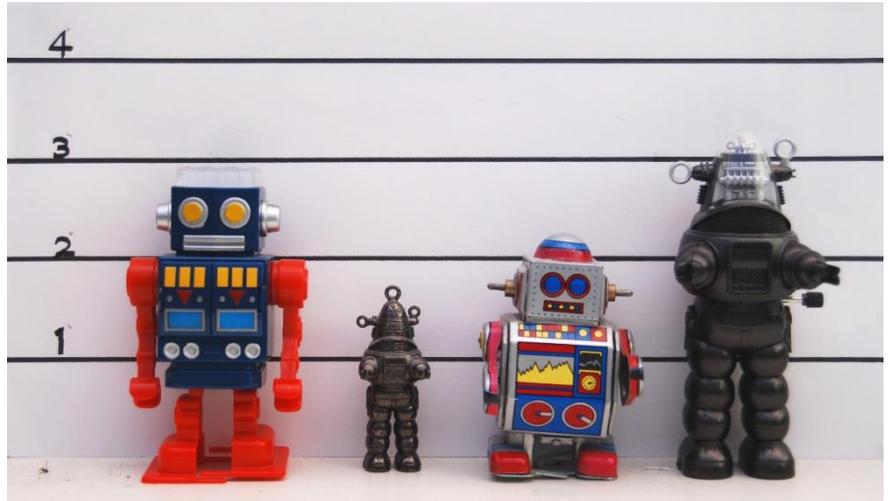
CJS'64

Aktuell haben es maschinelle Verfahren schwer in der GND



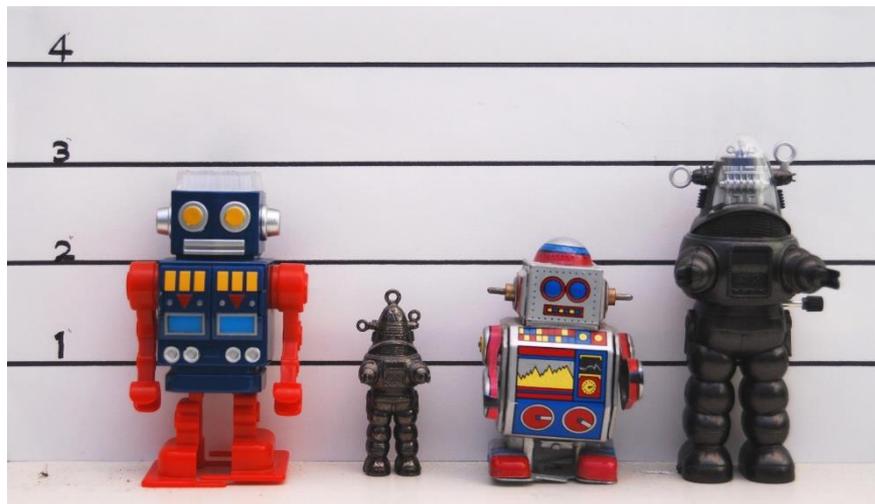
Maschinelle Verfahren autonom nur im Ausnahmefall einsetzbar

- Matching von Identifiern oder anderen eindeutigen Zeichenketten z.B. zum Ergänzen von Verknüpfungen
- Abgleichen von Dubletten mit großer Ähnlichkeit

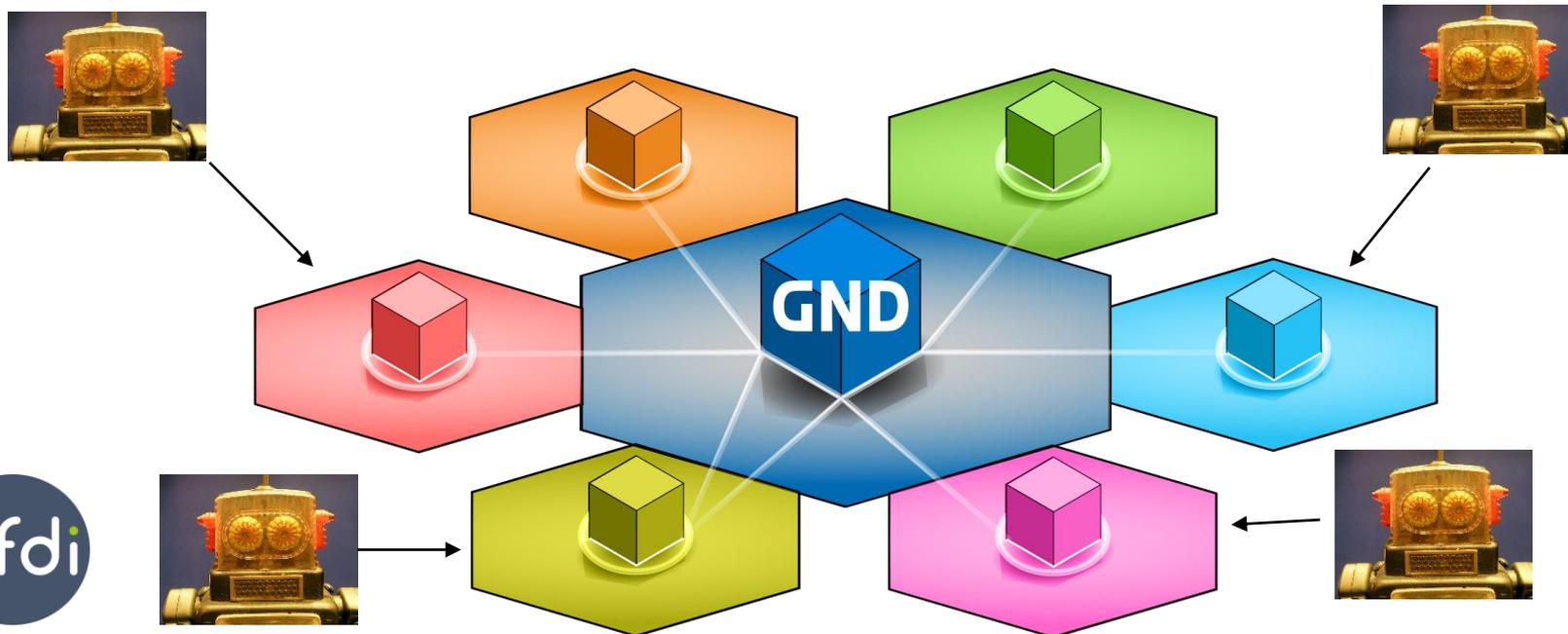


Gründe

- ! Können die Regeln nicht ausreichend einhalten
- ! Produzieren nach den bisherigen Maßstäben keine verlässliche Qualität (starke Streuung)
- ! Harmonisieren nur schwer mit der aktuellen Infrastruktur und den bestehenden Workflows



Das reicht aber nicht ...



Anforderung #1

Die GND muss auch in maschinellen Verfahren zum Einsatz kommen können – als Datenquelle und Ziel-Vokabular

Frage: Wie kann die Eignung der GND für den Einsatz in maschinellen Verfahren verbessert werden?

Anforderung #2

GND-Einträge entstehen meistens aus dem Erschließungsprozess heraus.

Frage: Wie finden neue benötigte Entitäten und Konzepte künftig Eingang in die GND, wenn die betroffenen Ressourcen ausschließlich maschinell erschlossen wurden?

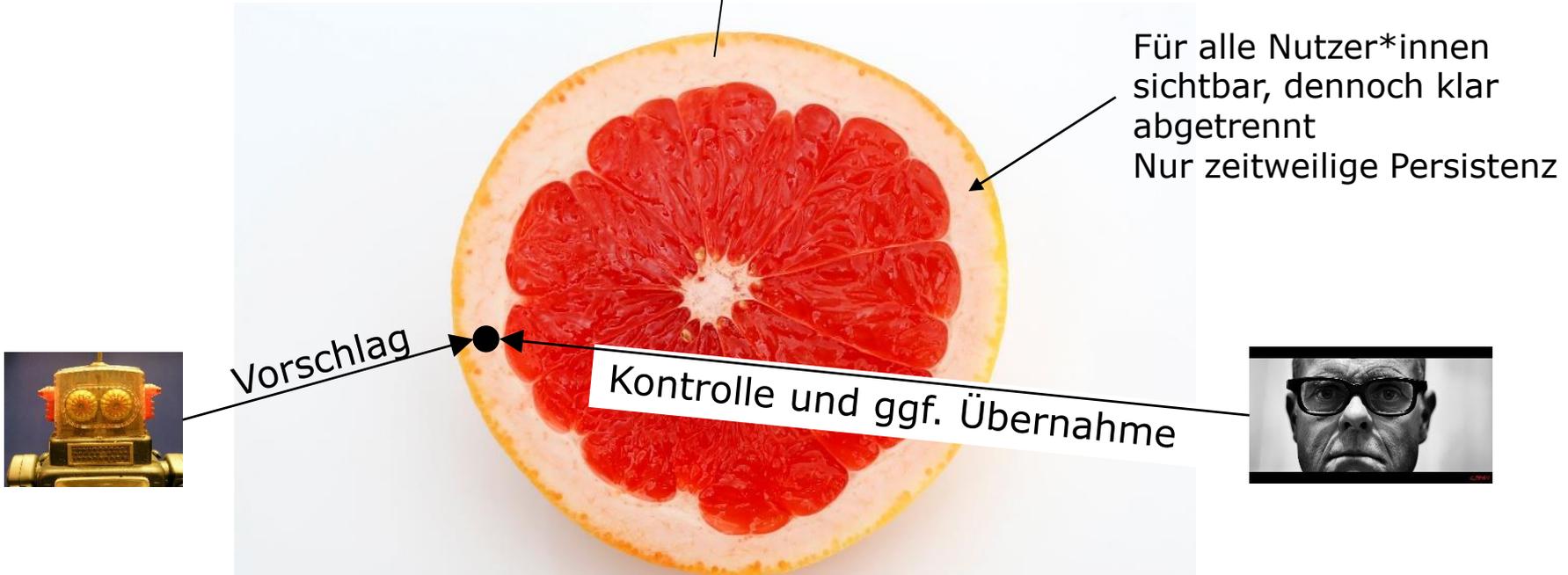
Anforderung #3

Man könnte maschinelle Verfahren auch zur Erleichterung der intellektuellen GND-Redaktion nutzen.

Frage: Welche Verbesserungen sind hier denkbar?

Thesen und Diskussion

These: Es wird eine Übergangszone benötigt



Nur für eine überschaubare Anzahl an Vorschlägen möglich!

Diskutierte Szenarien

- Vorschläge für neue Sachschlagwörter aus der automatischen Inhaltserschließung
 - Vorhaben in der DNB
- Vorschläge für Werktitel ermittelt durch Clustering-Verfahren basierend auf großen Metadatenbeständen wie Culturegraph, K10+
- Generell: Melden potentieller Fehler, Anomalien, Warnungen

Wunsch: Intelligente Vorschlagssteuerung



These: Es wird eine zusätzliche Datenschicht für maschinelle Verfahren benötigt

- Teilen der Data Provenance: Datenübernahme kann die maschinelle Herkunft verschleiern, Deep Fakes verhindern
- Teilen von Datenanreicherungen zur Unterstützung maschineller Verfahren bei der Kontextualisierung: Wortwolken, normalisierte Daten, ...
- Teilen von Analyseergebnissen: Hashes, Vektoren, Fehler und Warnungen, ...

These: Wir benötigen eine generelle über eine maschinen-freundliche GND

- Müssen Regeln angepasst werden und wie konkret?
- Wie muss sich die Datenhaltungs- und Schnittstellen-Infrastruktur weiterentwickeln?
- Wie müssen sich die Erschließungsoberflächen weiterentwickeln?

Vielen Dank!

Credits

- Folie 5: [Niels Wörheide](#): [Müngstener Brücke NRW.jpg](#)
- Folie 9: [Caesar Trinova](#): Figure-eight knot
- Folie 11: [Dan Dzurisin](#): Starling Murmuration Near Starved Rock State Park IL DDZ_0104
- Folie 12,13: [Azchael](#): Rütli Webstuhl mit Anbauautomat - 140 Jahre Zweigart - Tag der offenen Tür
- Folie 14: [Bennilover](#): Favorite Cinnamon Rolls
- Folie 15: [CJS*64](#): An Overnight intellectual
- Folie 17: [Michael Dain](#): robot
- Folie 21: [Tim Robinson](#): robots
- Folie 29: [Tim Reckmann](#): Grapefruit bzw. Pampelmuse