

Claudia Grote (DNB)

Werkstattgespräche: Prozesse und Modularisierung für die maschinelle Erschließung

Erschließungsprozesse mit Textanalyse

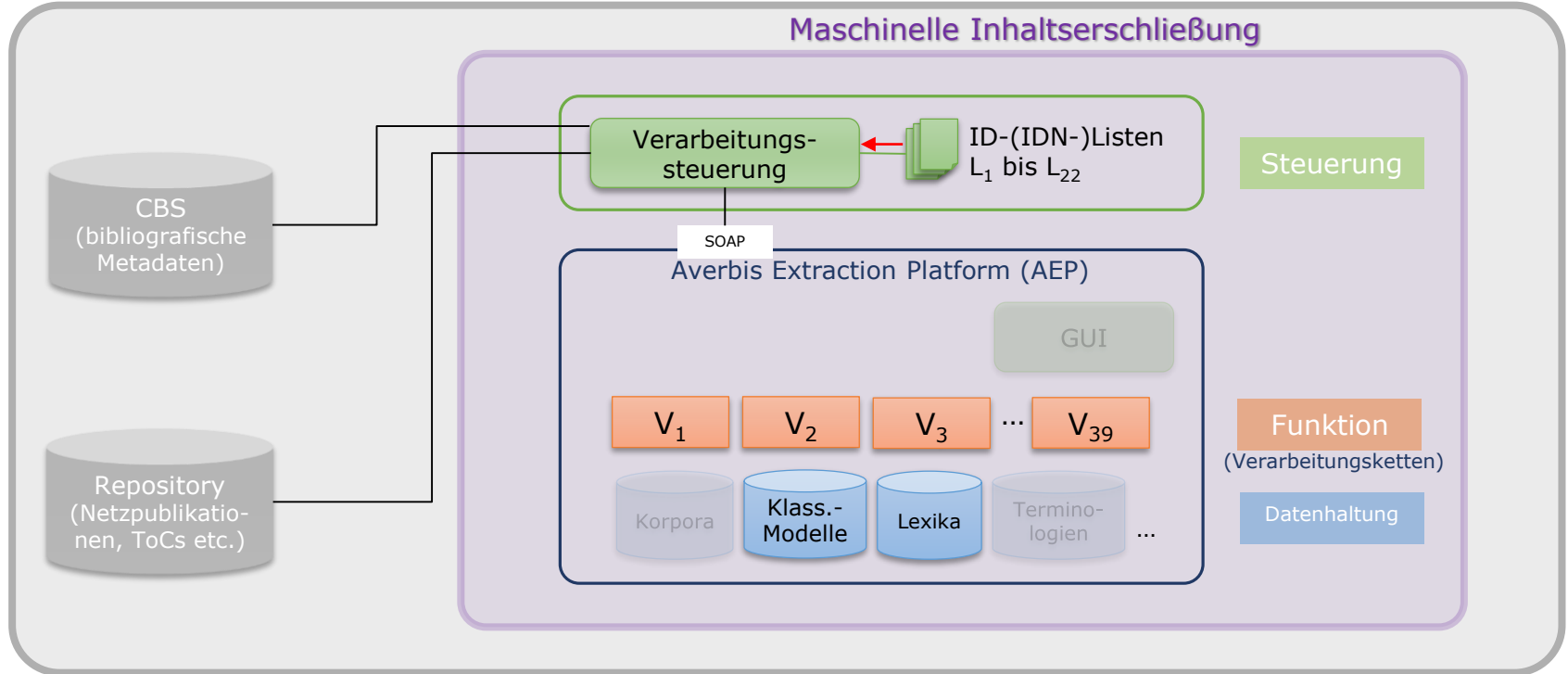
39 produktive Verarbeitungsketten für 22 Funktionen der Vergabe von

- Sprachencode nach ISO 639-1
- DNB-Sachgruppen
- GND-Schlagwörtern
 - deutschsprachige Artikel
 - deutschsprachige BOD-Verlagspubl.
- Schlagwörtern aus GND- und LCSH-Vok.
 - Inhaltsverzeichnisse
 - wissenschaftliche Arbeiten
- DDC-Kurznotationen
 - Informatik (004)
 - Sozialwissenschaften, Soziologie, Anthropologie (300)
 - Naturwissenschaften (500)
 - Mathematik (510)
- DDC-Kurznotationen (Forts.)
 - Astronomie, Kartographie (520)
 - Physik (530)
 - Chemie (540)
 - Technik (600)
 - Medizin, Gesundheit (610)
 - Elektrotechnik, Elektronik (621.3)
 - Ingenieurbau und Umwelttechnik (624)
 - Hauswirtschaft und Familienleben (640)
 - Management (650)
 - Technische Chemie (660)
 - Industrielle und handwerkliche Fertigung (670)
 - Hausbau, Bauhandwerk (690)

IST-Architektur Maschinelle Inhaltserschließung

DNB-Infrastruktur

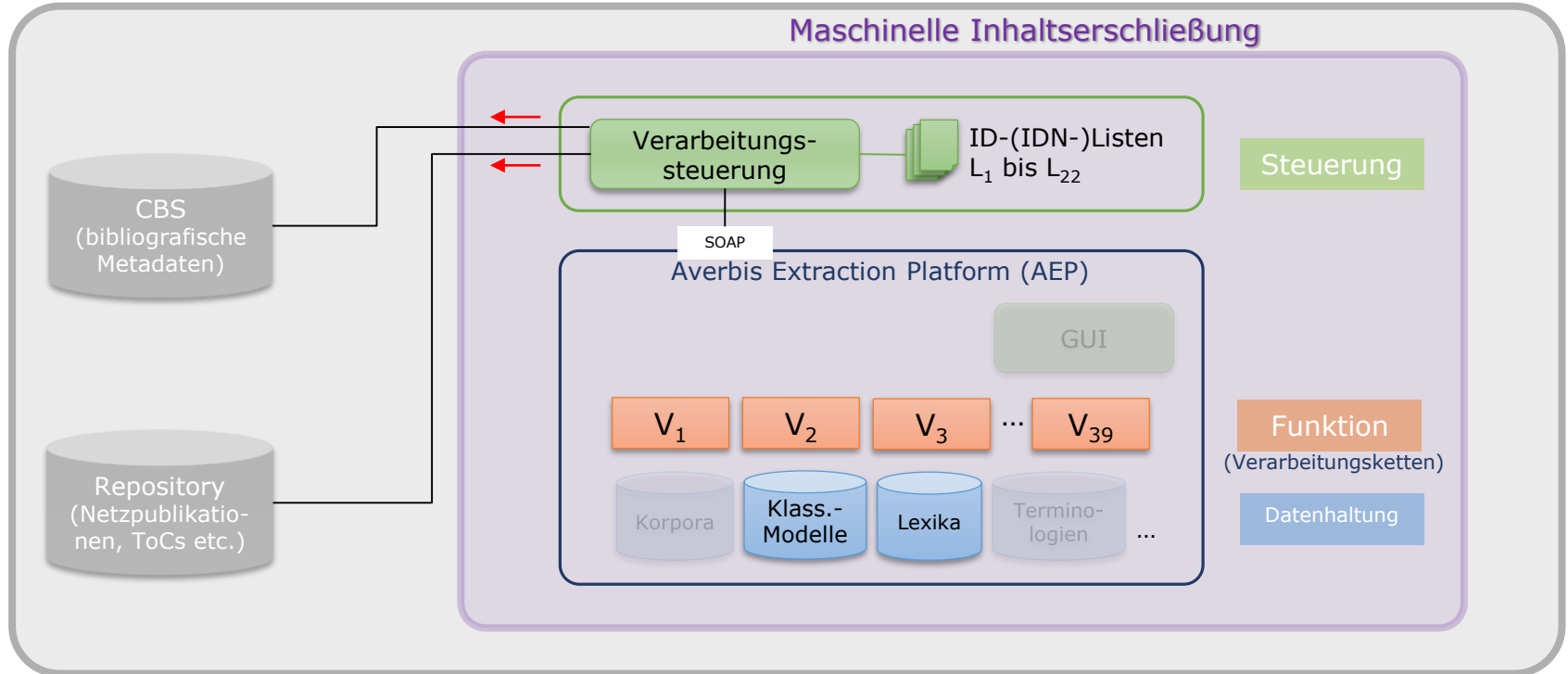
Maschinelle Inhaltserschließung



IST-Architektur Maschinelle Inhaltserschließung

DNB-Infrastruktur

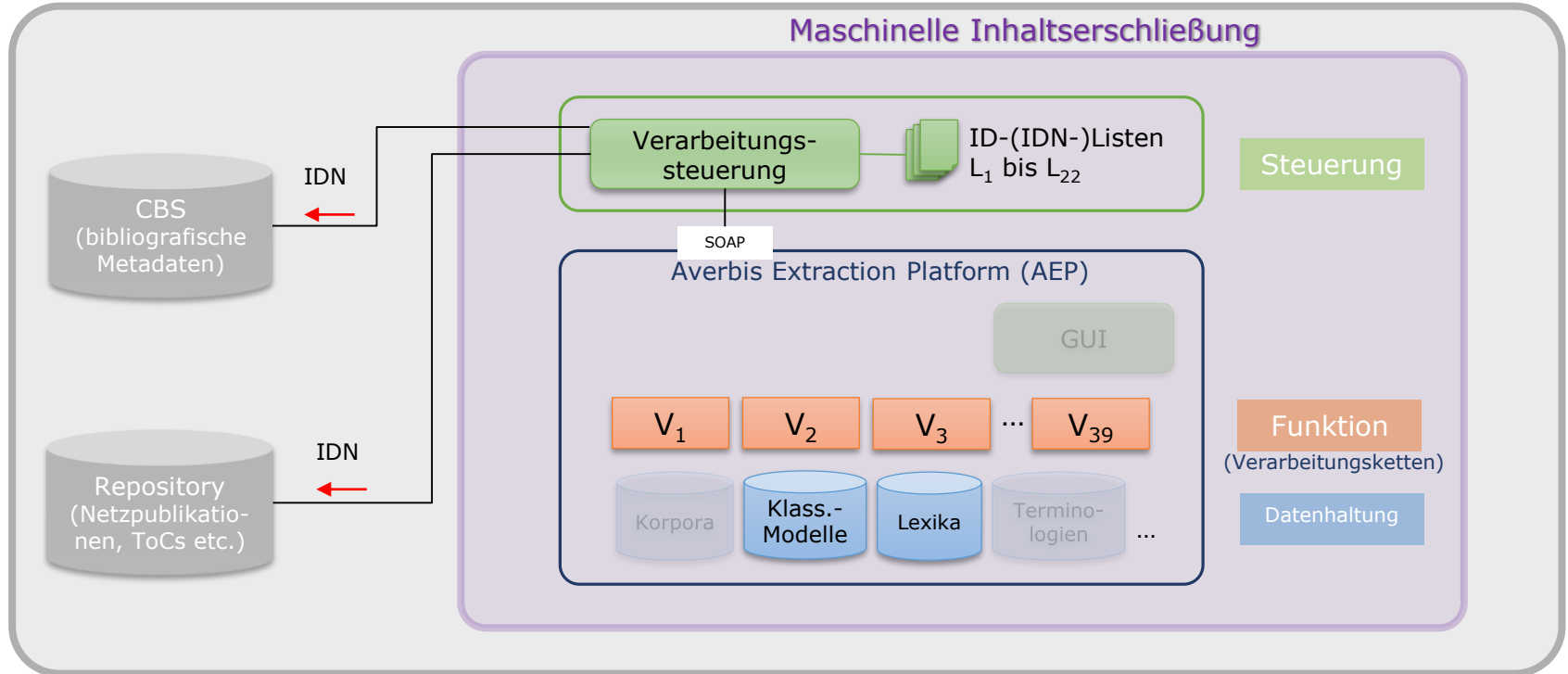
Maschinelle Inhaltserschließung



IST-Architektur Maschinelle Inhaltserschließung

DNB-Infrastruktur

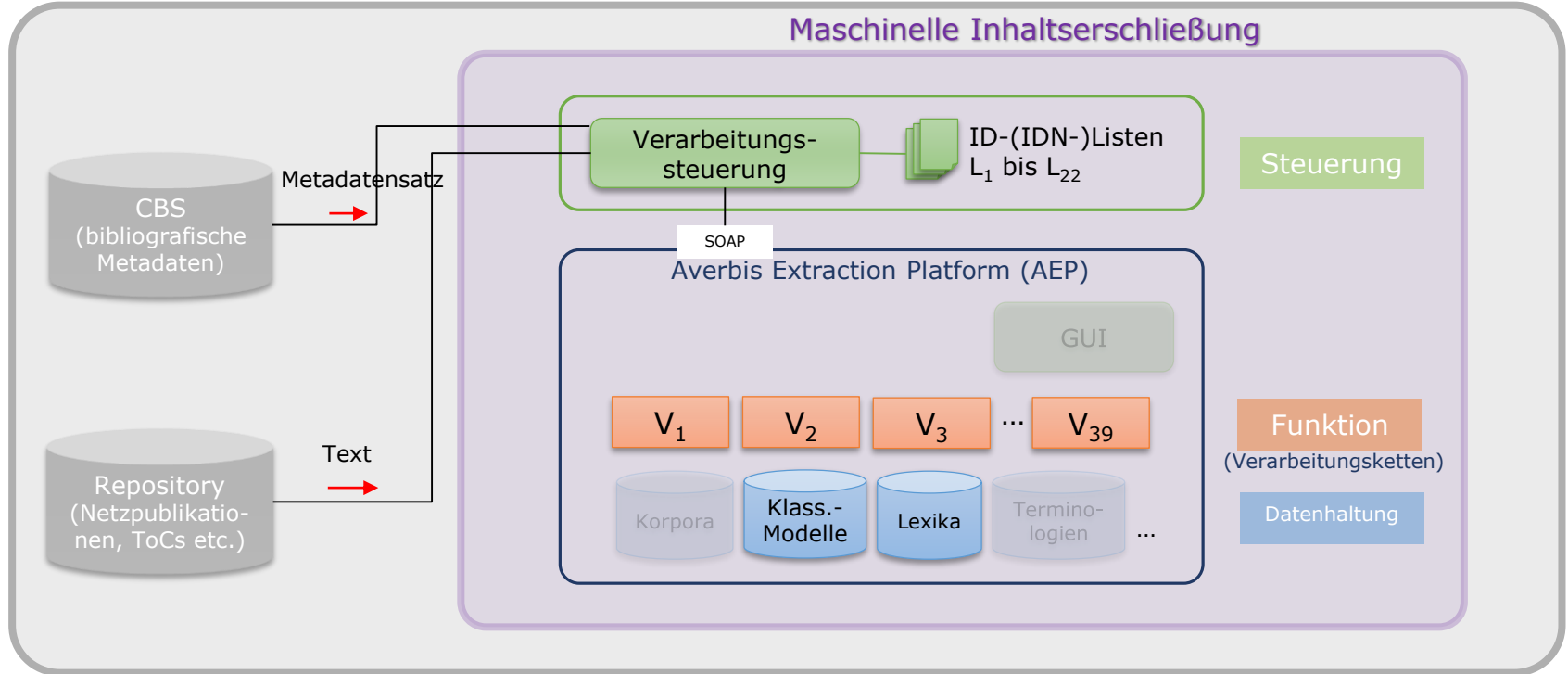
Maschinelle Inhaltserschließung



IST-Architektur Maschinelle Inhaltserschließung

DNB-Infrastruktur

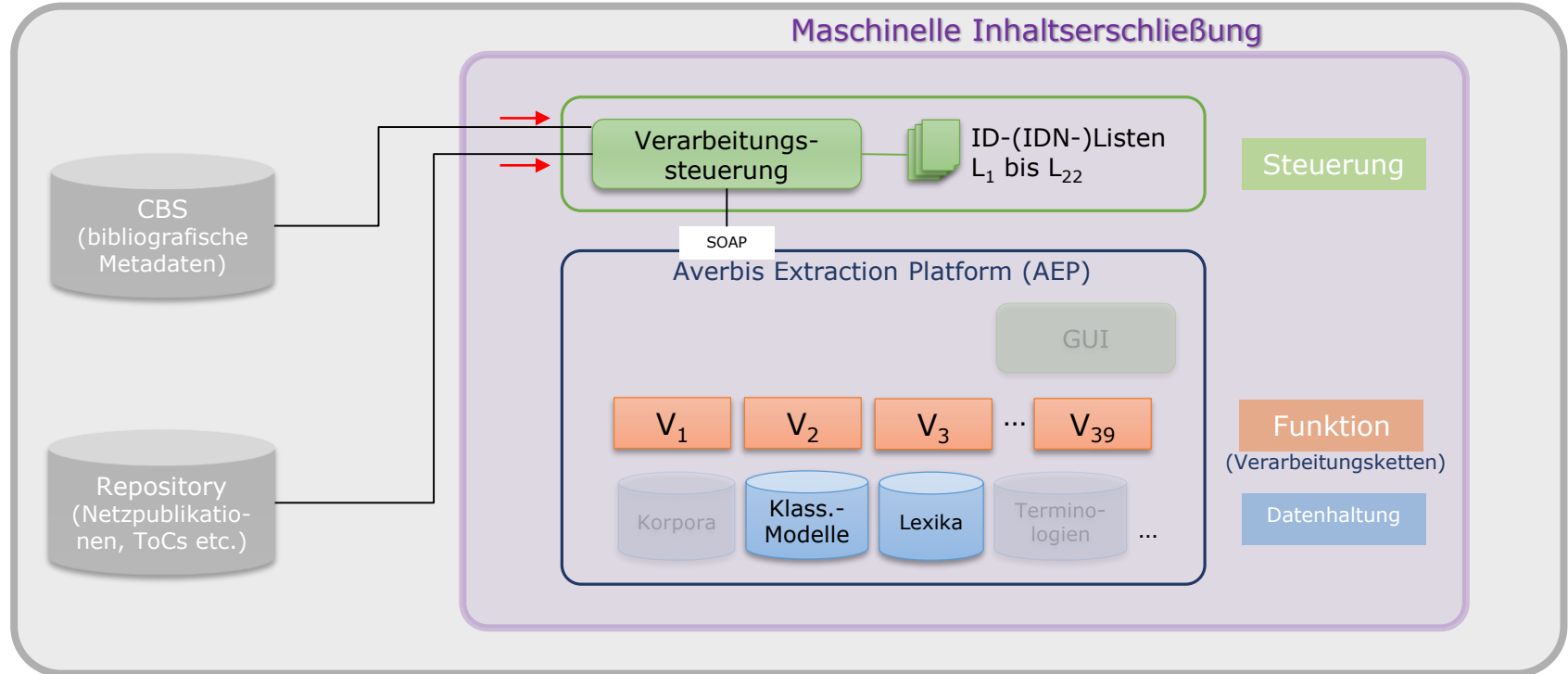
Maschinelle Inhaltserschließung



IST-Architektur Maschinelle Inhaltserschließung

DNB-Infrastruktur

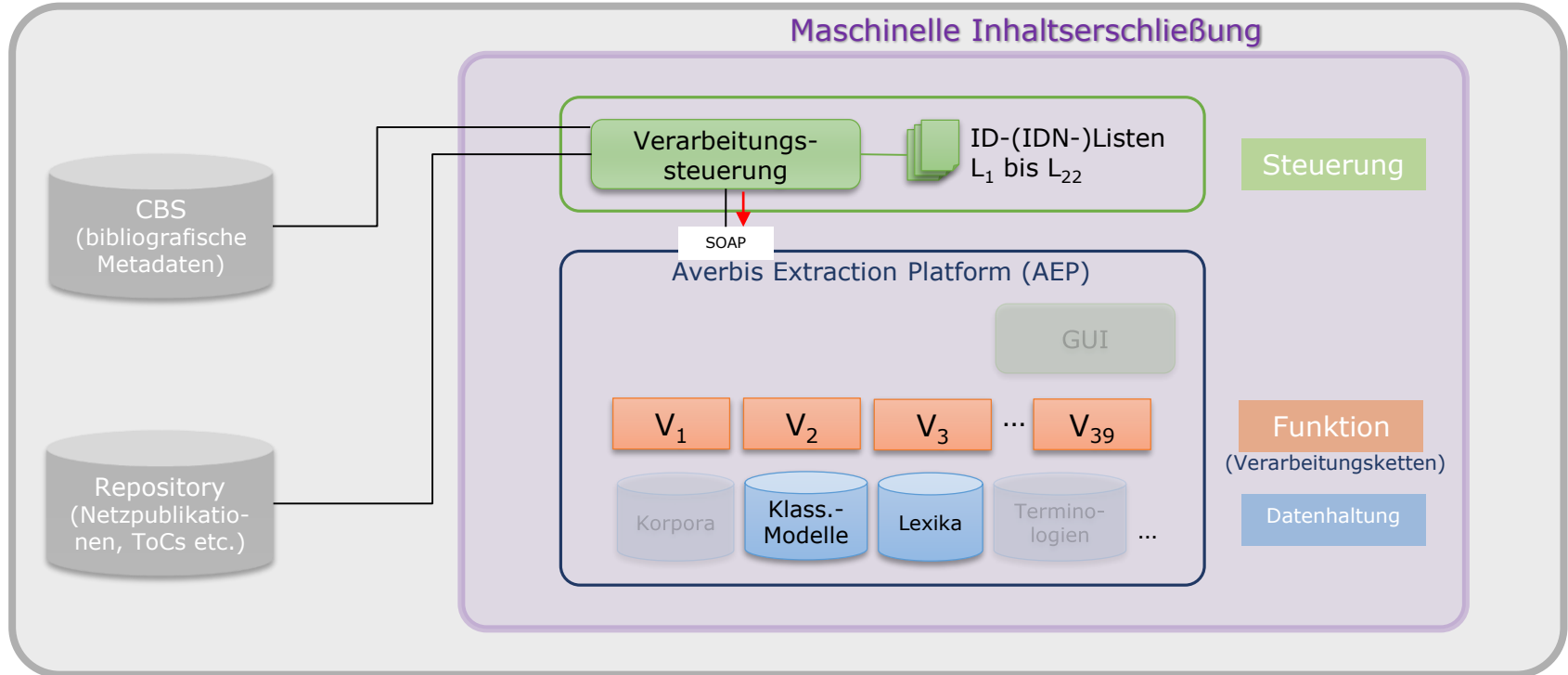
Maschinelle Inhaltserschließung



IST-Architektur Maschinelle Inhaltserschließung

DNB-Infrastruktur

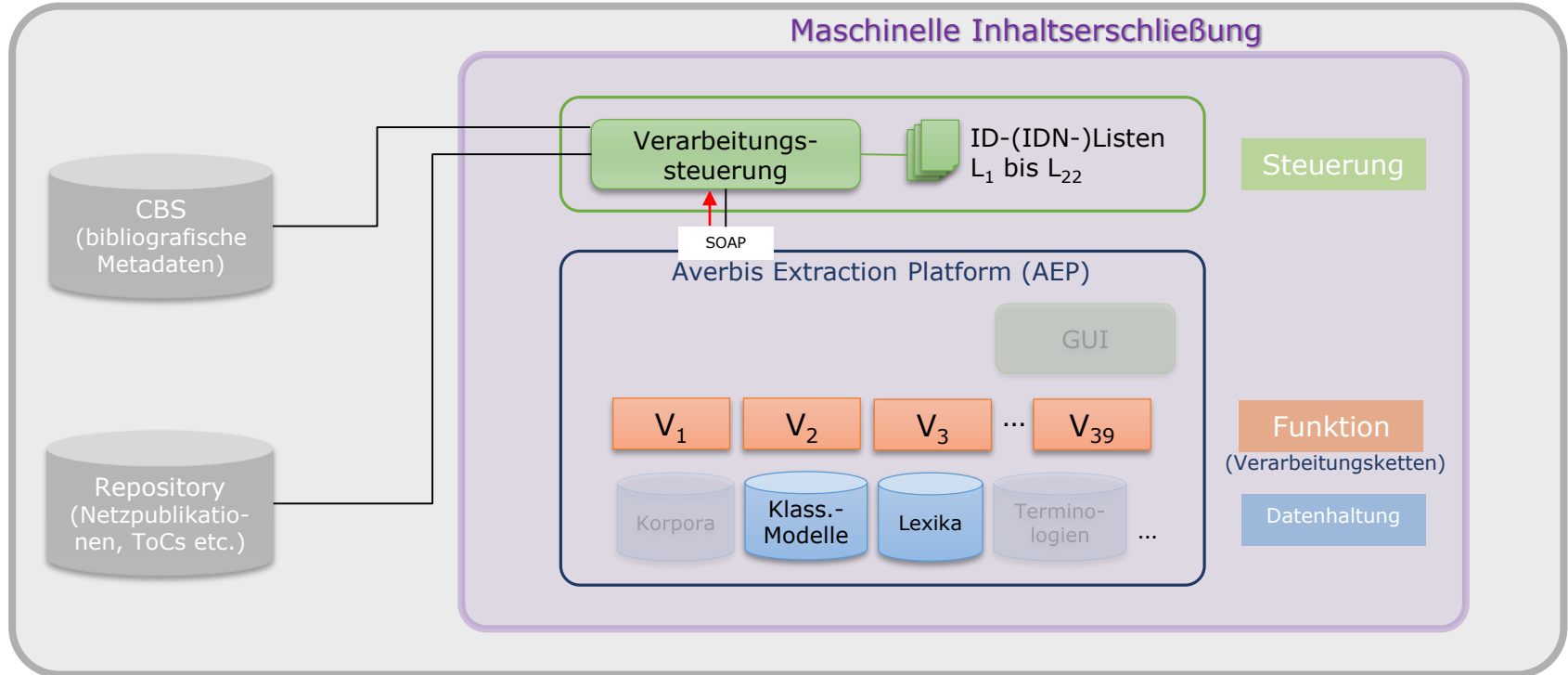
Maschinelle Inhaltserschließung



IST-Architektur Maschinelle Inhaltserschließung

DNB-Infrastruktur

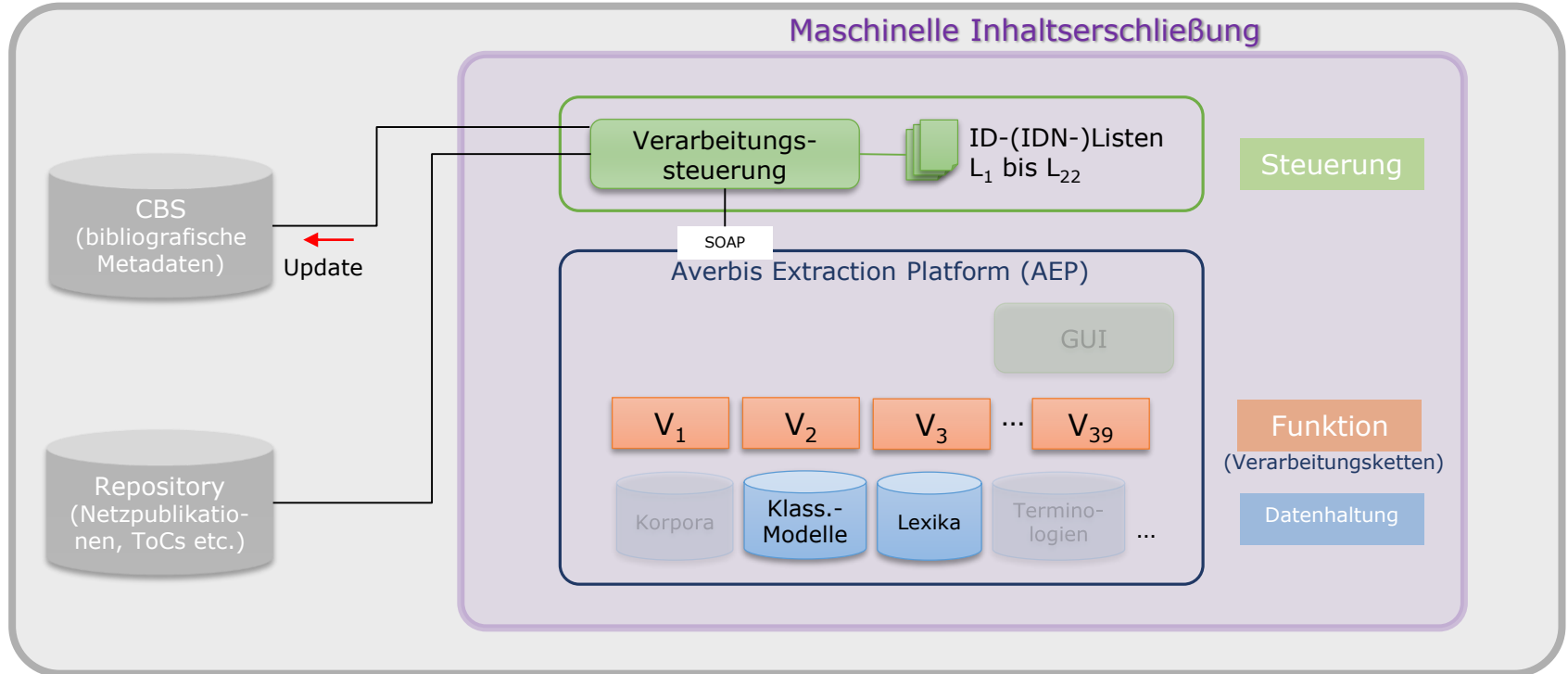
Maschinelle Inhaltserschließung



IST-Architektur Maschinelle Inhaltserschließung

DNB-Infrastruktur

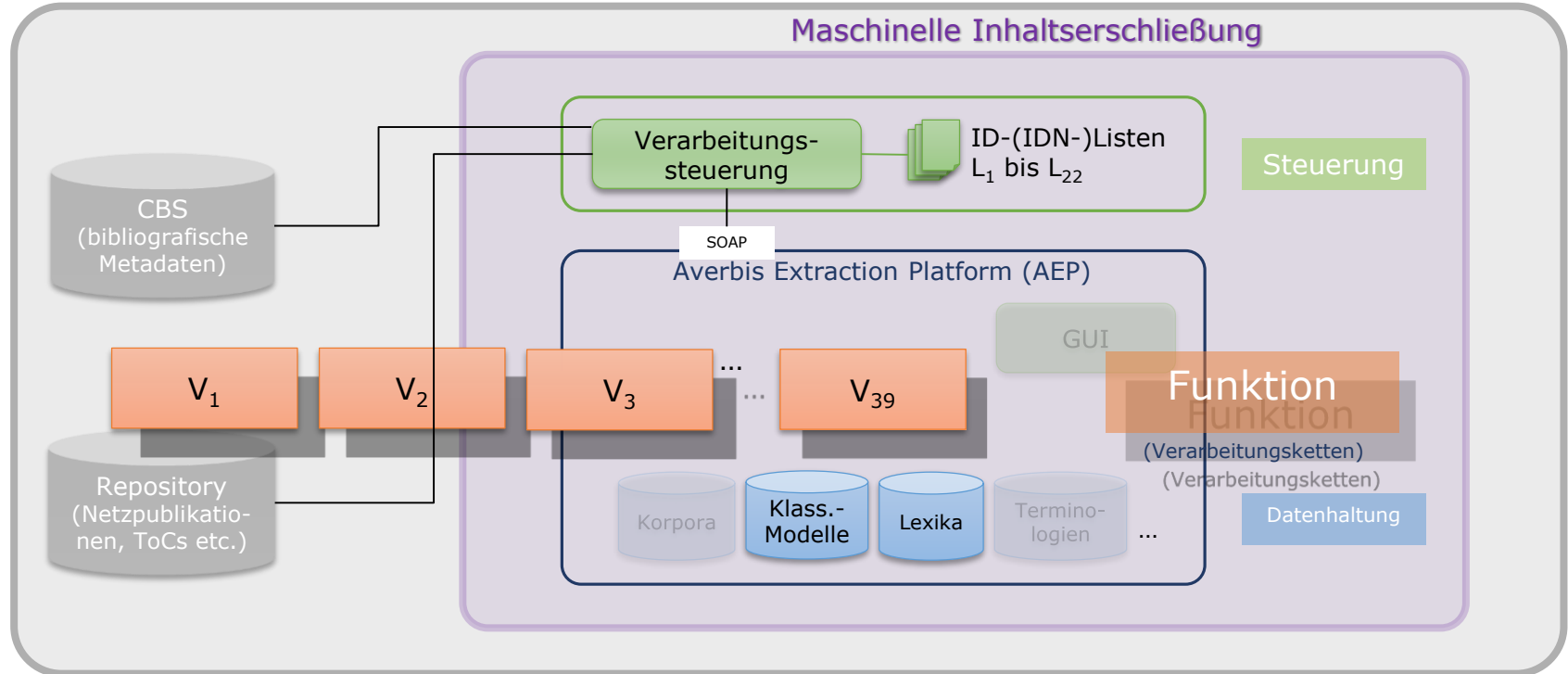
Maschinelle Inhaltserschließung



IST-Architektur Maschinelle Inhaltserschließung

DNB-Infrastruktur

Maschinelle Inhaltserschließung



Von Verarbeitungsketten zu modularer Verarbeitung

V₁

**Klassifikation
DNB-Sachgruppen**

Textvorverarbeitung



Textnormalisierung



Merkmalberechnung



Klassifikation

V₂

**Beschlagwortung
mit GND-Vokabular**

Textvorverarbeitung



Textnormalisierung



Wörterbuch-Abgleich
und Disambiguierung



Termgewichtung
und -auswahl

V₃

**Klassifikation
DDC Kurznotation 004**

Textvorverarbeitung



Textnormalisierung



Merkmalberechnung



Klassifikation

...

V₃₉

**Klassifikation
DDC Kurznotation 300**

Textvorverarbeitung



Textnormalisierung

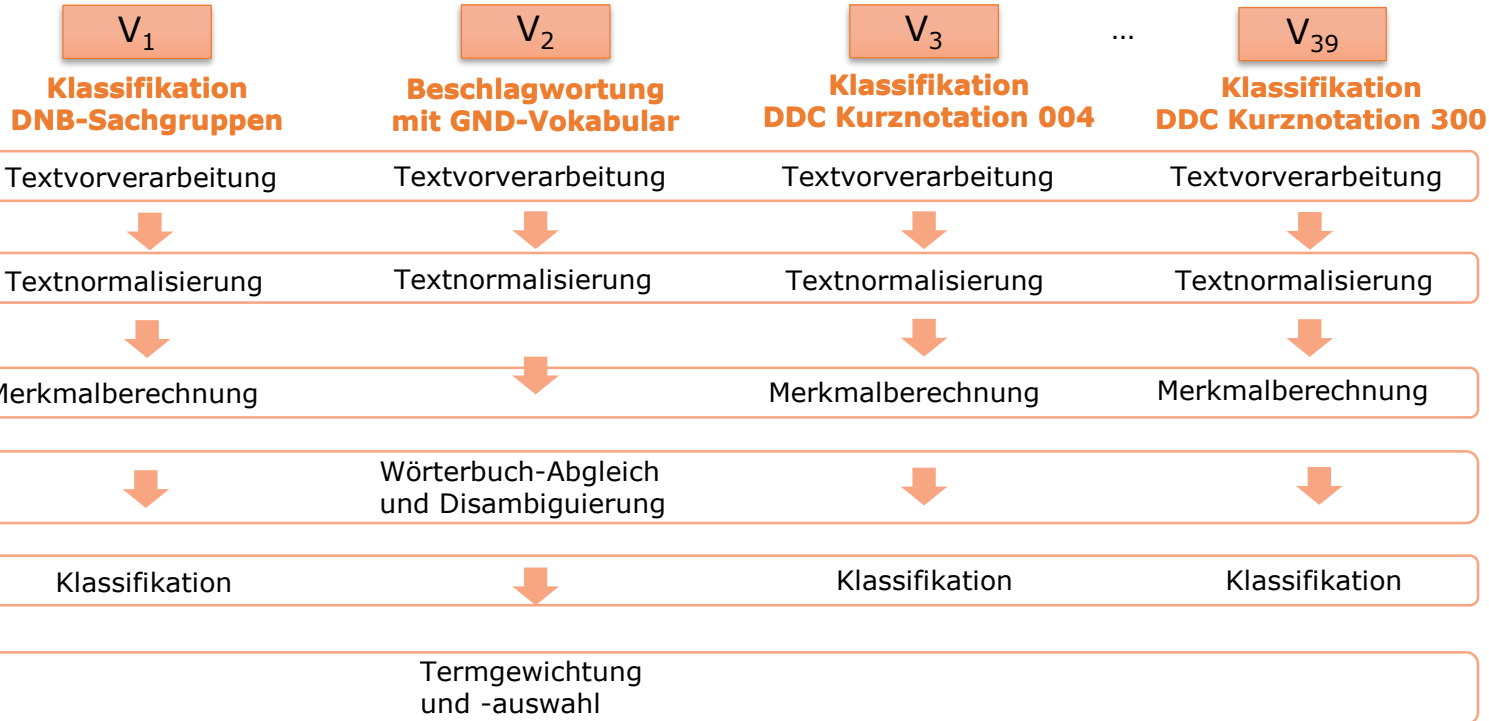


Merkmalberechnung



Klassifikation

Von Verarbeitungsketten zu modularer Verarbeitung



Von Verarbeitungsketten zu modularer Verarbeitung

Textvorverarbeitung

Textvorverarbeitung

Textvorverarbeitung

Textvorverarbeitung

Modul₁

Textnormalisierung

Textnormalisierung

Textnormalisierung

Textnormalisierung

Modul₂

Merkmalberechnung

Merkmalberechnung

Merkmalberechnung

Modul₃

Wörterbuch-Abgleich
und Disambiguierung

Modul₄

Klassifikation

Klassifikation

Klassifikation

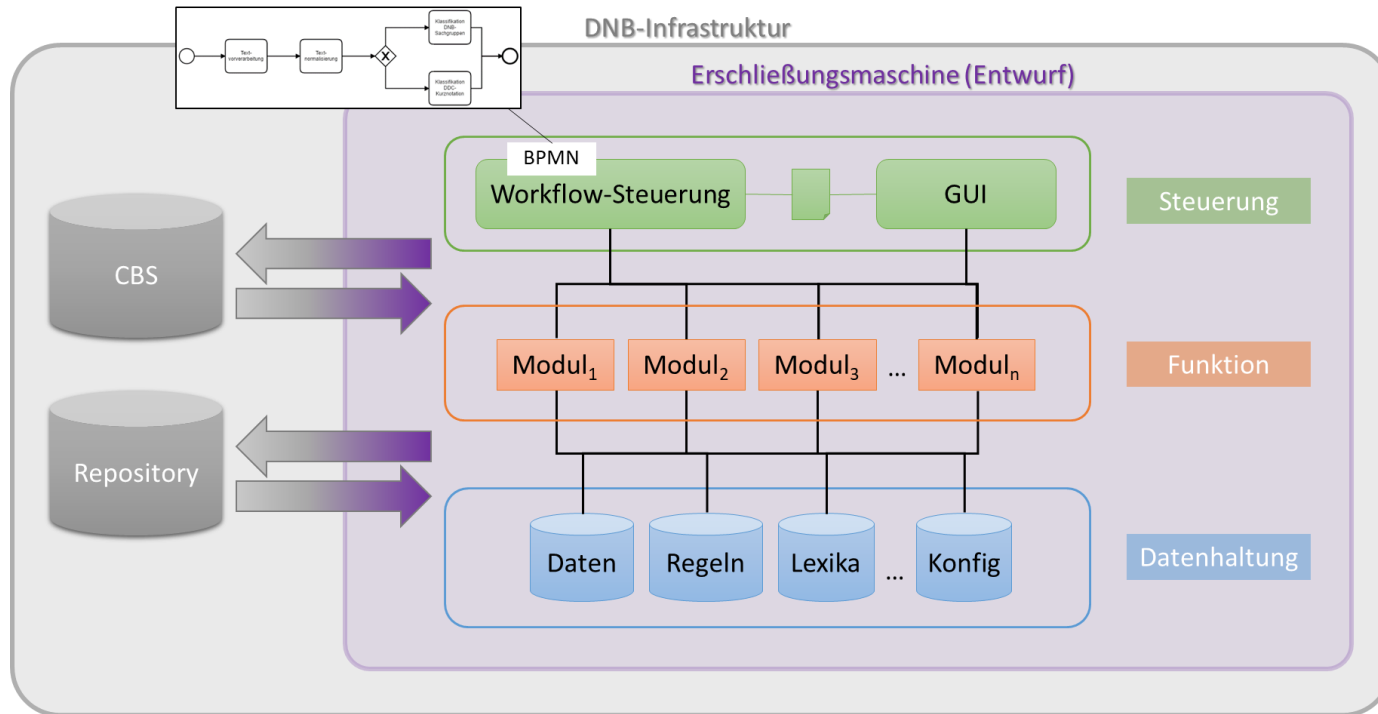
Modul₅

Termgewichtung
und -auswahl

Modul₆

⋮

SOLL-Architektur Maschinelle Inhaltserschließung (Entwurf)



Kontakt:

Claudia Grote

IT / Data-Workflow und Geschäftsprozesse

Tel. 069 – 1525 1782

c.grote@dnb.de