

A-Tag 2016

Wien – 16. September 2016



PDF/UA Real Life Check 2016

Eine Studie zum Stand der PDF-Barrierefreiheit auf den Internetseiten österreichischer Bundesministerien (inklusive Bundeskanzleramt)



I. Ein paar Worte vorweg

- Was wollten wir wissen?
- Studiendesign
- Welche Limitierungen haben wir bewusst in Kauf genommen?

1 | Was wollten wir wissen?

- Wie ist der Stand der PDF-Barrierefreiheit (Menge und Qualität barrierefreier PDF-Dokumente)?
- Ist PDF/UA bei den österreichischen Ministerien angekommen?
- Ist PDF/UA erreichbar?
- Welches sind die häufigsten PDF/UA-Probleme?
- Wie lassen sich große Mengen an PDF-Dokumenten schnell und effizient prüfen („Monitoring“)?
- Wie lassen sich Dokumente mit massiven Mängeln („schwarze Schafe“) schnell identifizieren?
- Wie lassen sich Erfolge im Bereich PDF-Barrierefreiheit messbar machen und kommunizieren (nach innen und nach außen)?
- Basis für Vergleiche zwischen Organisationen und Stand zu einem bestimmten Zeitpunkt

2 | Studiendesign

- 13 Bundesministerien inkl. Bundeskanzleramt:
<https://www.bka.gv.at/site/3353/default.aspx>
- Zeitraum: 1 Jahr: Alle PDF-Dateien, die Google Crawler für den Zeitraum 01.07.2015 bis 30.06.2016 findet
- Prüfkriterien: Alle Maschinen prüfbaren PDF/UA-Anforderungen gemäß Matterhorn-Protokoll
- Heuristische Auswertungen, um Indizien für eine ungeeignete Semantik zu finden (erste Versuche)
- Stichprobenmäßige Detail- und Plausibilitätsprüfungen zur Sicherstellung der Datenqualität
- Alle Dokumente sind dokumentiert mit Downloaddatum, URL und Testergebnis

3 | Limitierungen

- Schnelles Prüfen großer Dokumentmengen muss sich auf automatische Prüfungen fokussieren
- Google Crawler findet nicht alle tatsächlich veröffentlichten PDF-Dokumente im entsprechenden Zeitraum
- Dokumente mit Parsingfehlern wurden übersprungen
- Keine Berücksichtigung der textlichen Kennzeichnung nicht barrierefreier PDF-Dokumente
- Kein Überprüfen, ob Inhalte in einer alternativen zugänglichen Form vorliegen (beispielsweise als HTML)



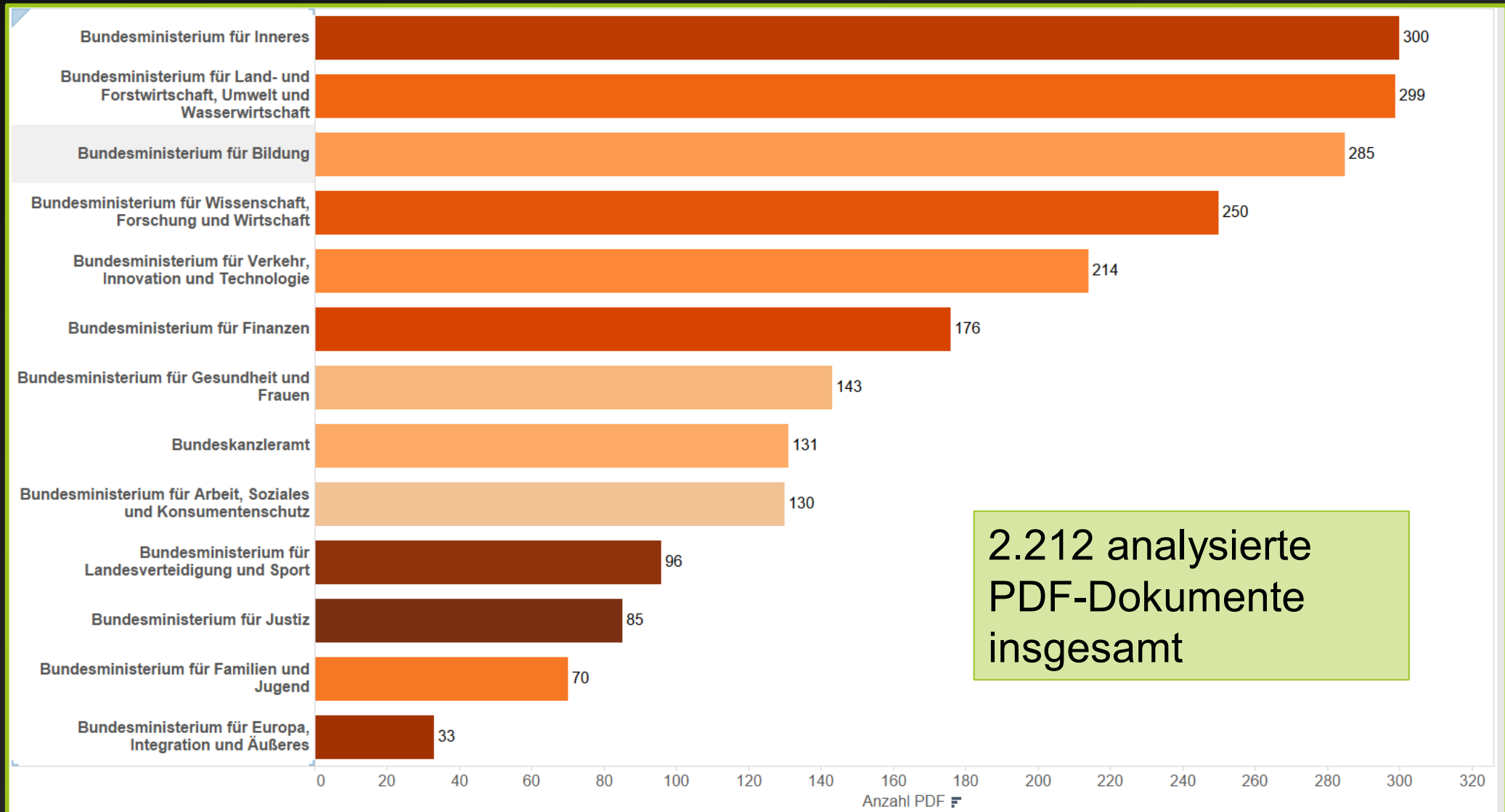
II. Ergebnisse

1. Grundlegende Zahlen
2. PDF-Dokumente mit Tags
3. PDF/UA-Qualität gemäß Maschinen-prüfbarer Anforderungen
4. Einzelne PDF/UA-Anforderungen:
Welche Probleme treten am häufigsten auf?
5. Indizien für „Semantic Richness“ (Heuristische Auswertung)
6. 100% PDF/UA

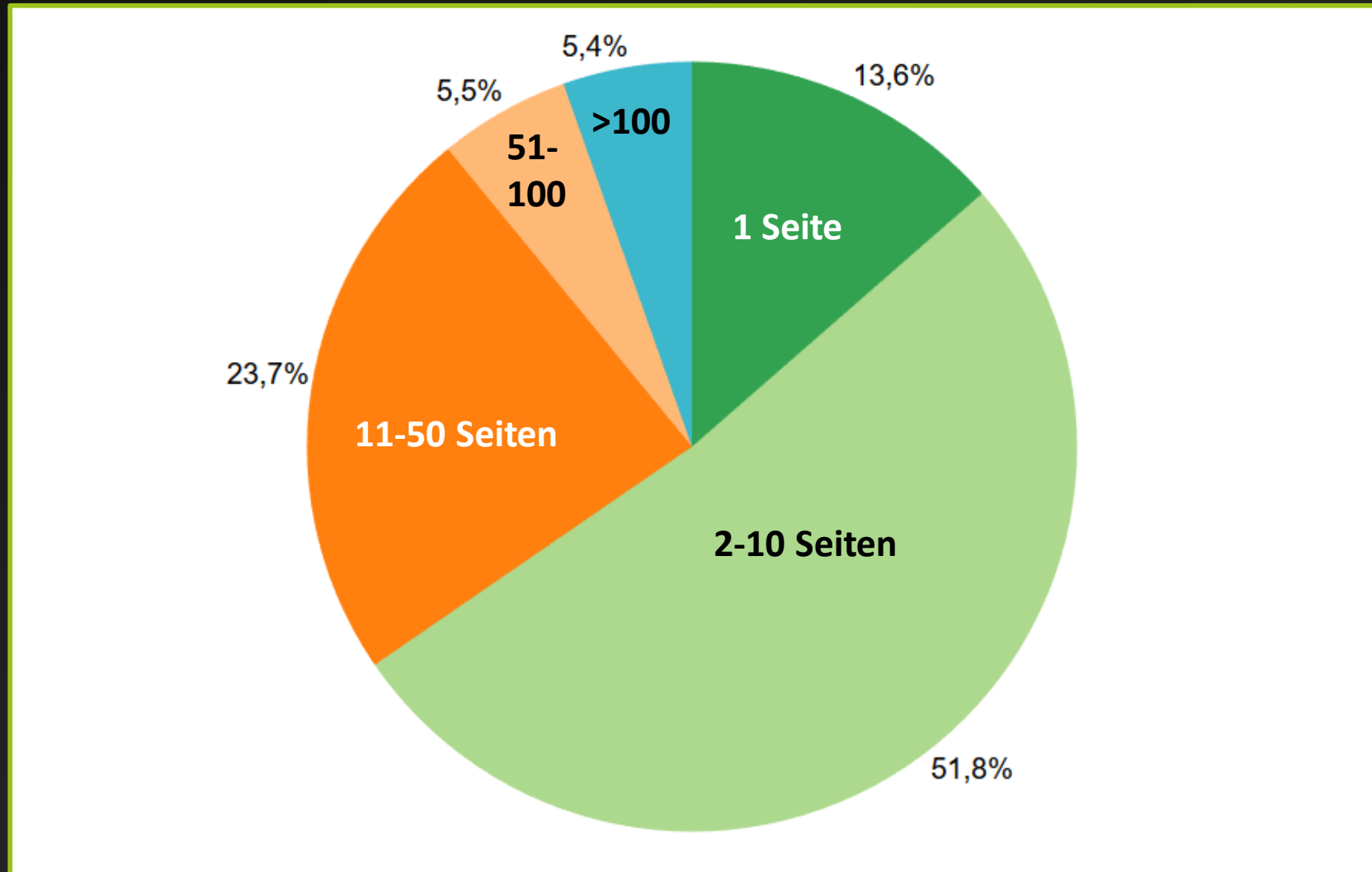
1 | Grundlegende Zahlen

- Anzahl der analysierten Dokumente
- Dokumentkategorien nach Seitenumfang
- Grundsprachen

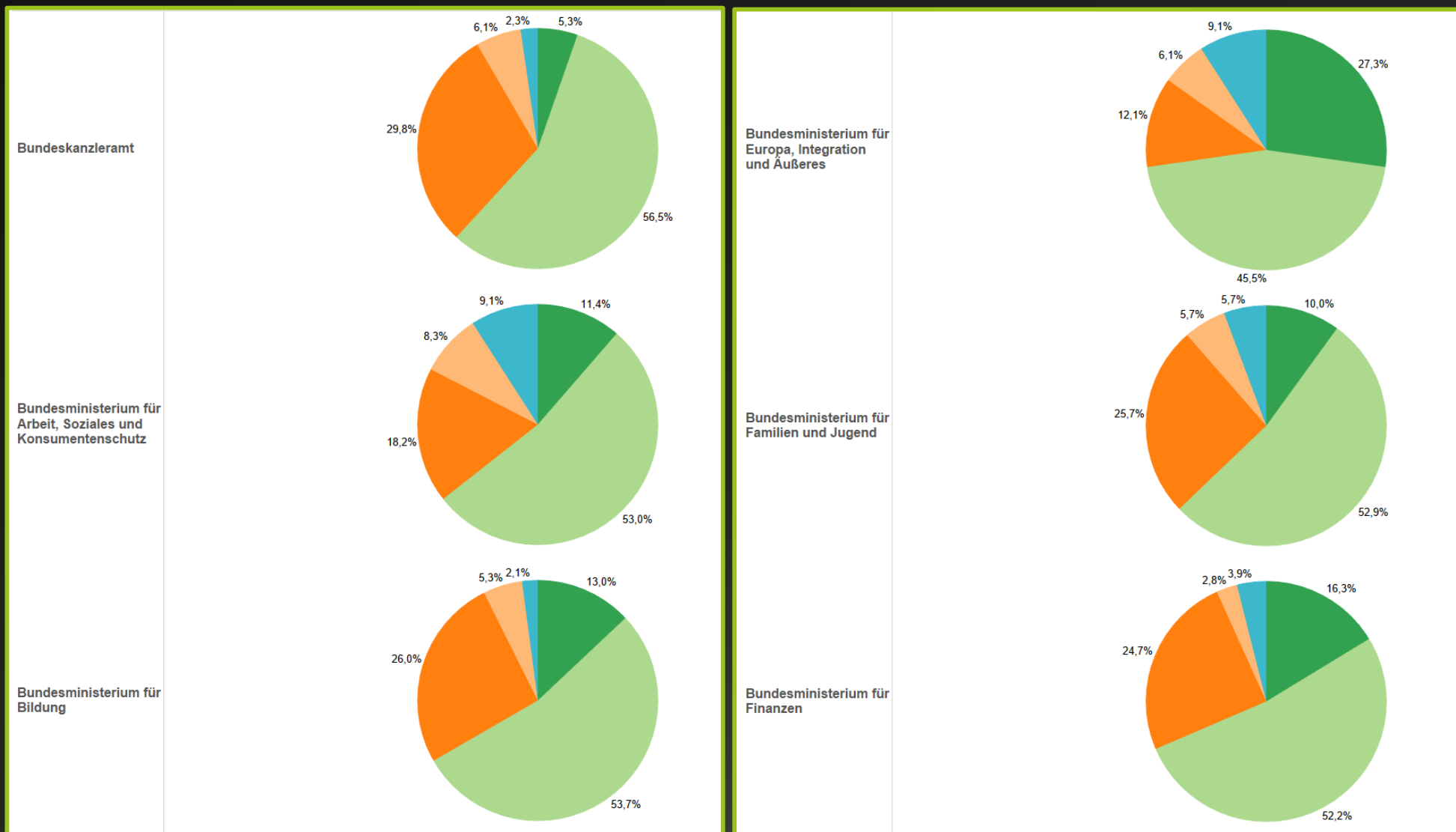
a | Anzahl der analysierten Dokumente



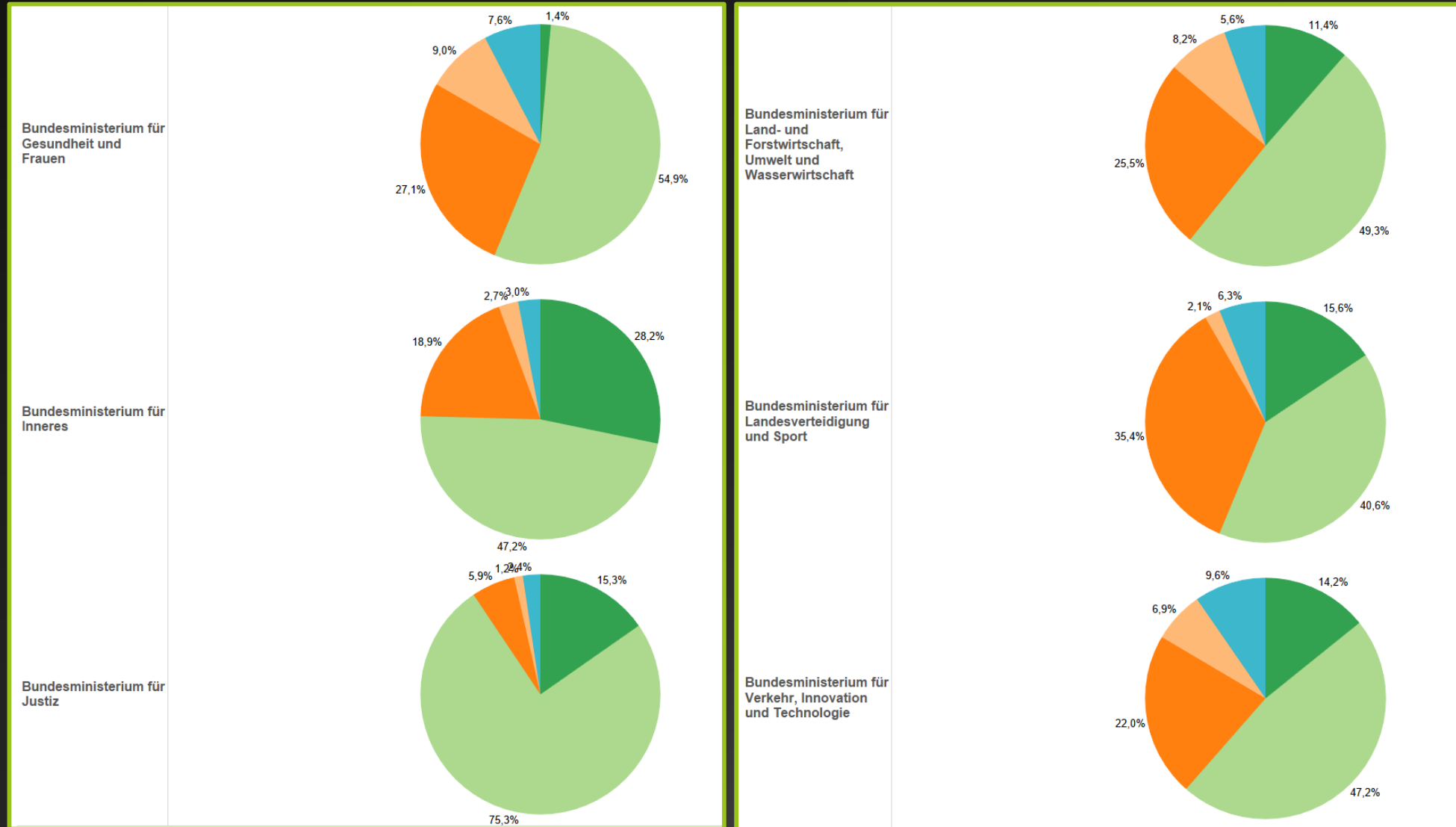
b | Dokumentkategorien nach Seitenumfang



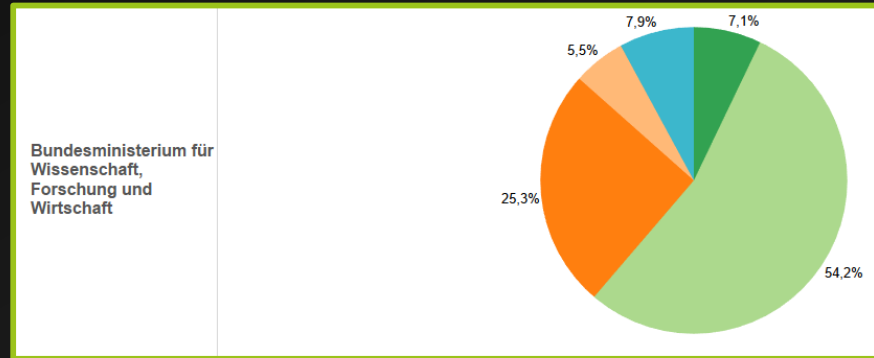
b | Dokumentkategorien (einzelne Ministerien)



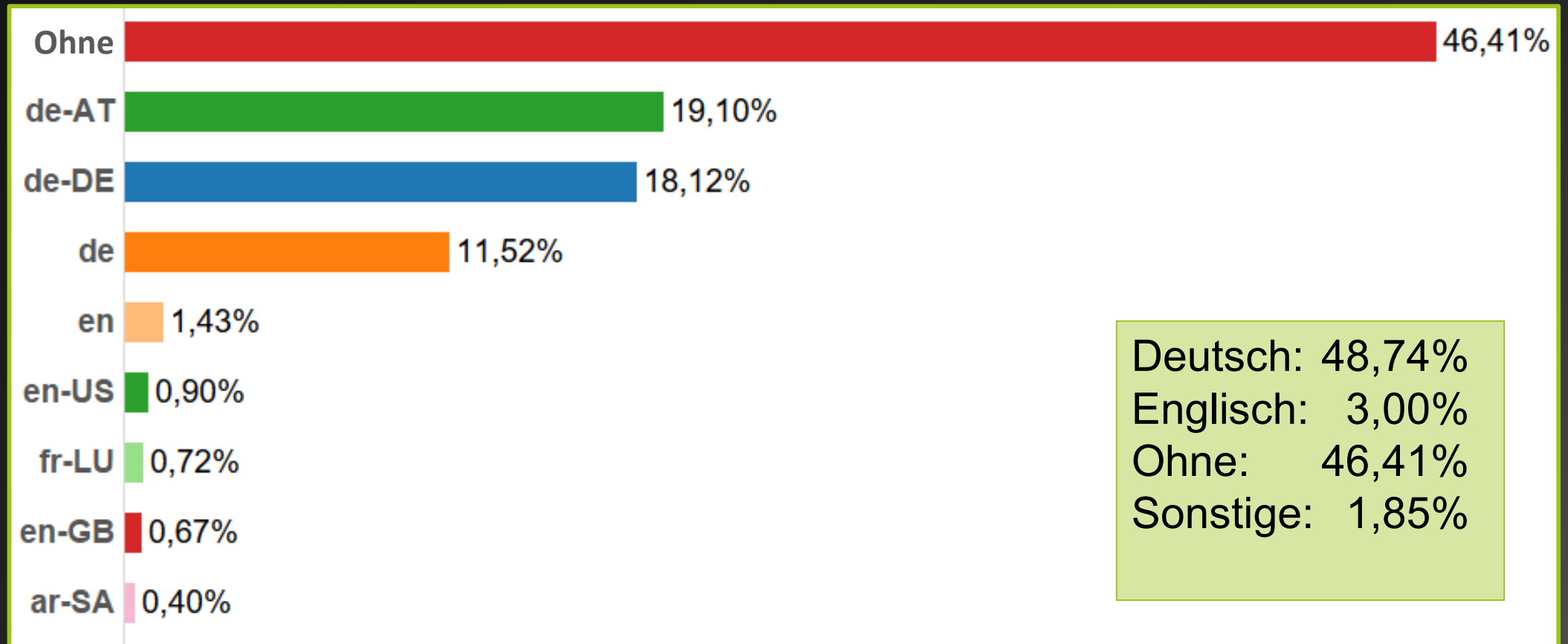
b | Dokumentkategorien (einzelne Ministerien)



b | Dokumentkategorien (einzelne Ministerien)



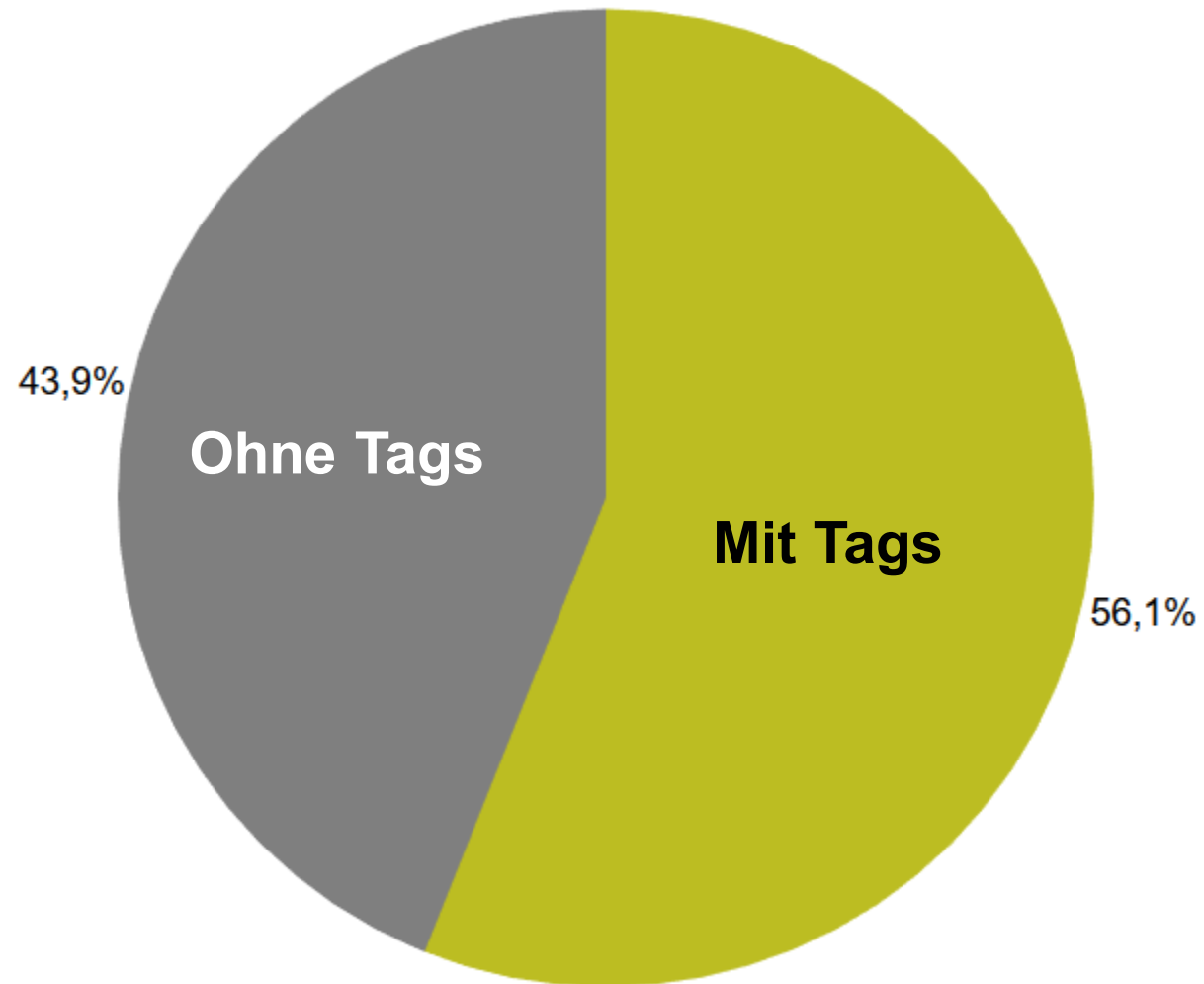
c | Grundsprachen



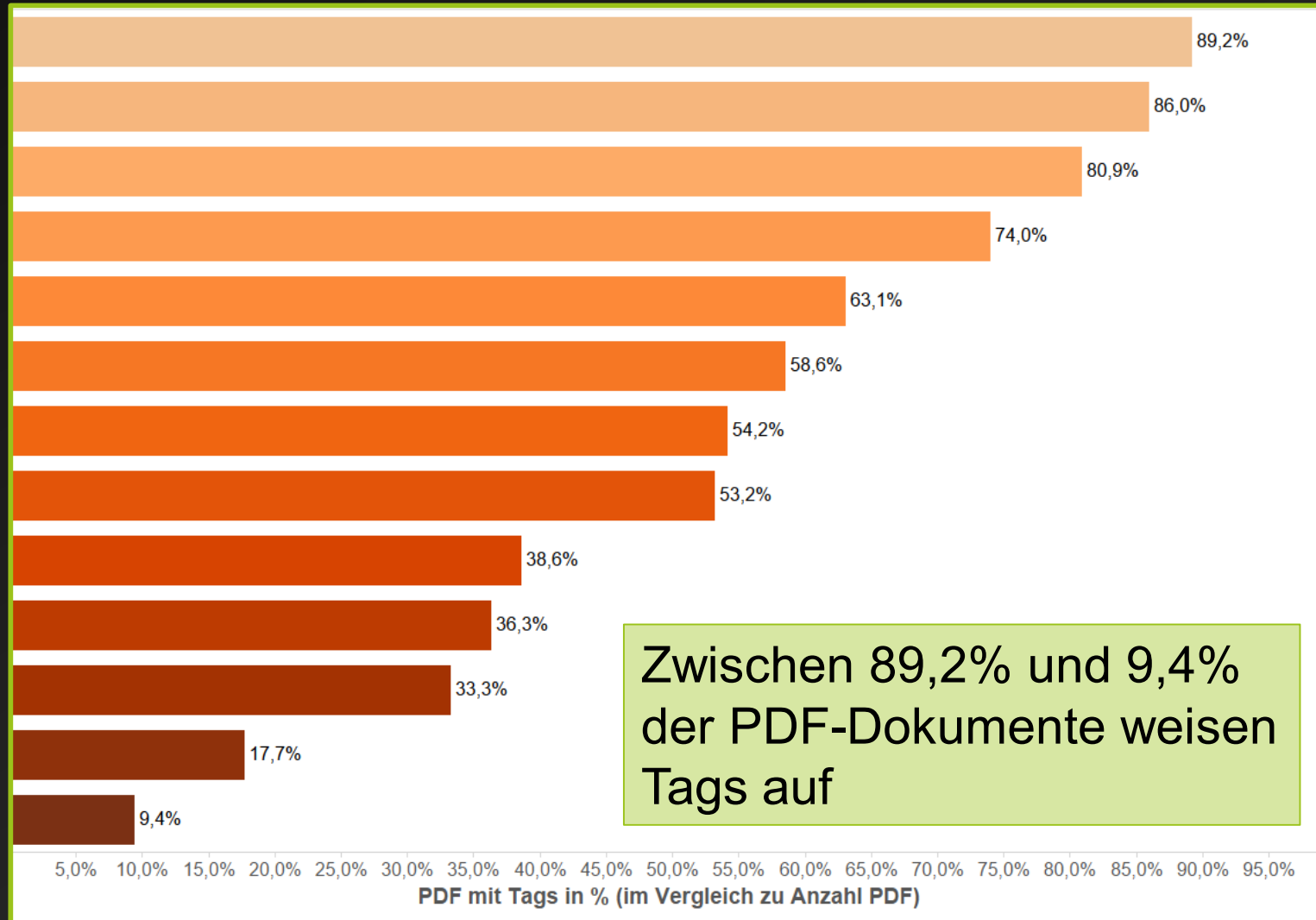
2 | PDF-Dokumente mit Tags

- Wieviele Dokumente weisen bereits Tags auf?
(Gesamttendenz)
- Unterschiede zwischen den Ministerien

a | PDF mit Tags (Gesamttendenz)



b | PDF mit Tags (einzelne Ministerien)



3 | PDF/UA-Qualität

- Wie verstehen wir PDF/UA-Qualität im Rahmen dieser Studie?
- Wie haben wir sie berechnet?
- Wie ist die Kennzahl einzuschätzen?
- PDF/UA-Qualität bei den einzelnen Ministerien
- **Gesamttendenz**

a | PDF/UA-Qualität: Konzept

- Bezieht sich nur auf **Maschinen-prüfbare PDF/UA-Anforderungen**
- **Dynamischer Maximalwert** als 100% (abhängig vom Dokument), da nicht alle Prüfpunkte für jedes Dokument relevant sind
- **Algorithmus** (pro Dokument und pro Prüfpunkt)

b | PDF/UA-Qualität: Berechnung

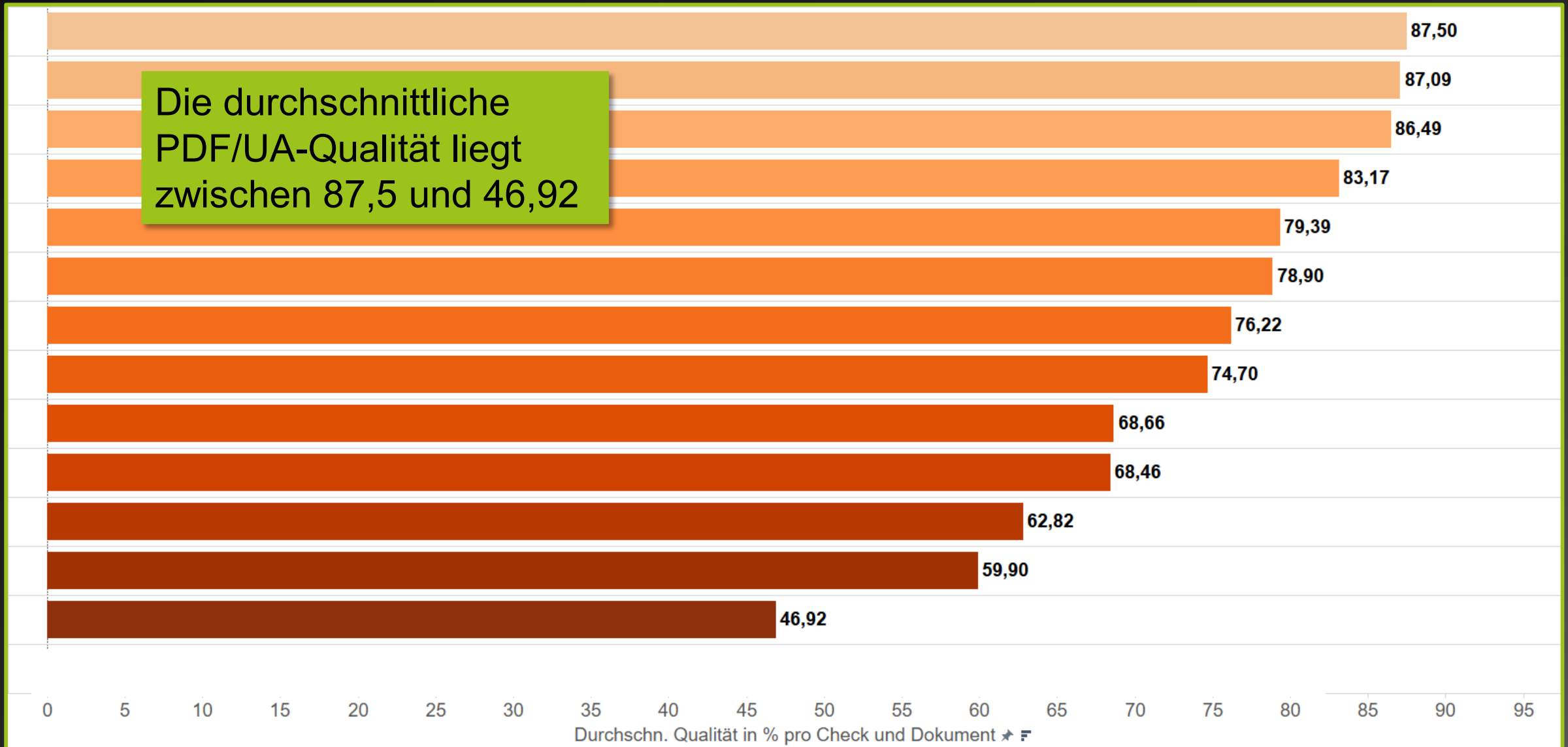
- Wieviel % eines einzelnen Prüfpunktes sind erfüllt
(dieser kann aus einem Check oder aus vielen Checks bestehen)
- **Gewichtung eines Prüfpunktes** für das Gesamtergebnis pro Dokument gemäß Schwere der Barriere im Hinblick auf WCAG 2-Kriterien
- Die Kennzahl gibt den **Restanteil nach Abzug der Strafpunkte in %** an (bezogen auf den Dynamischen Maximalwert)

c | Kennzahl für die PDF/UA-Qualität

Einschätzung

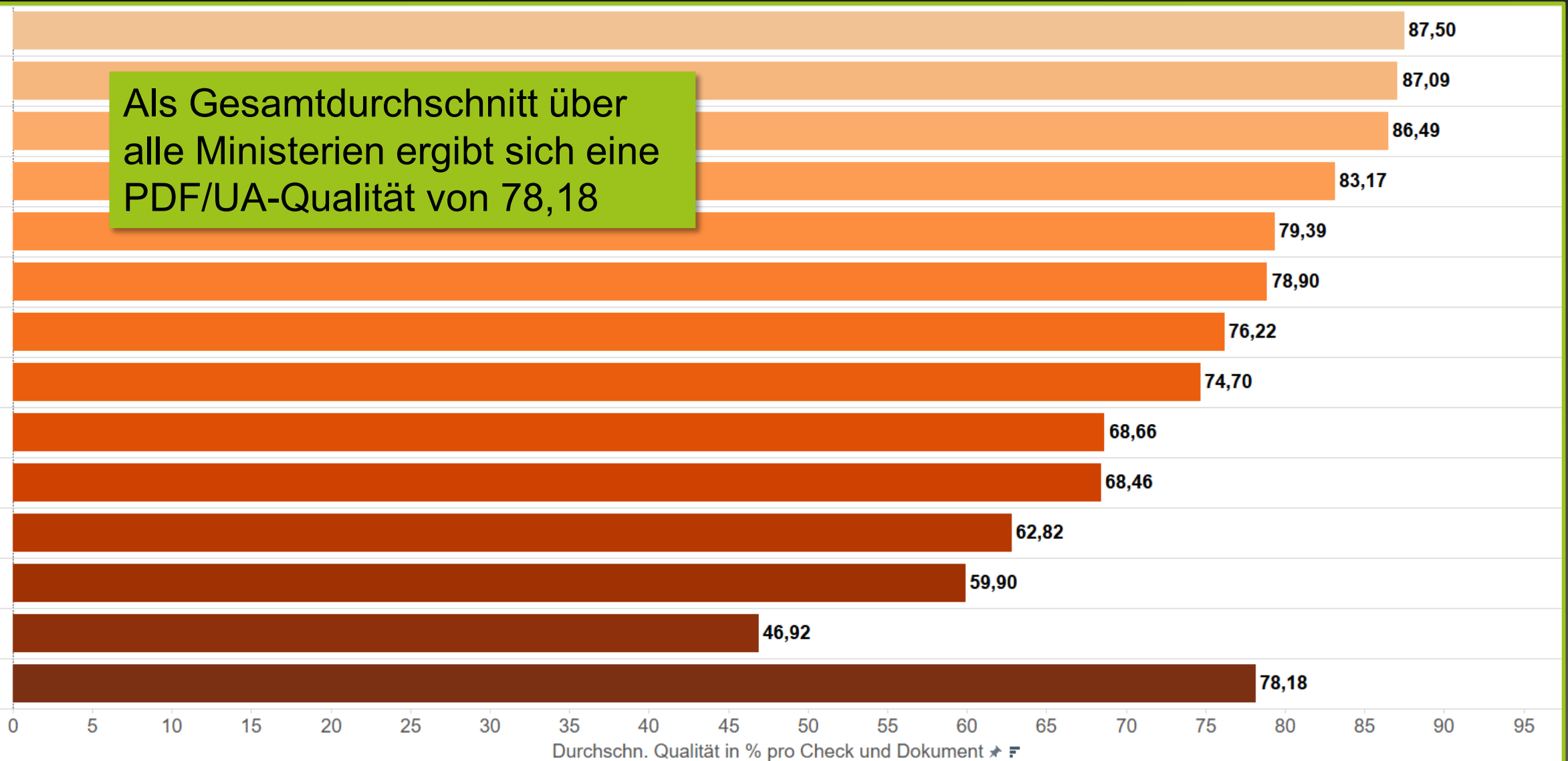
- Letztlich sind 100% das Ziel:
Es ist ja nicht die finale Einschätzung der Barrierefreiheit, sondern nur deren technische, Maschinen-prüfbare Basis
- unter 95 können die Dokumente nicht als barrierefrei bezeichnet werden
- unter 90% weist das Dokument gravierende Barrieren auf: Übersetzt heißt 90%, dass 10% der Informationen verloren gehen.

d | PDF/UA-Qualität (nach Ministerien)



d | PDF/UA-Qualität (Gesamttendenz)

Als Gesamtdurchschnitt über alle Ministerien ergibt sich eine PDF/UA-Qualität von 78,18



d | Was bedeutet 80% Qualität?

Stellen wir uns eine Hängebrücke mit 100 Latten vor ...



Diese entsprechen 100%.

d | Was bedeutet 80% Qualität?

80% bedeutet, dass jede 5. Latte fehlt.



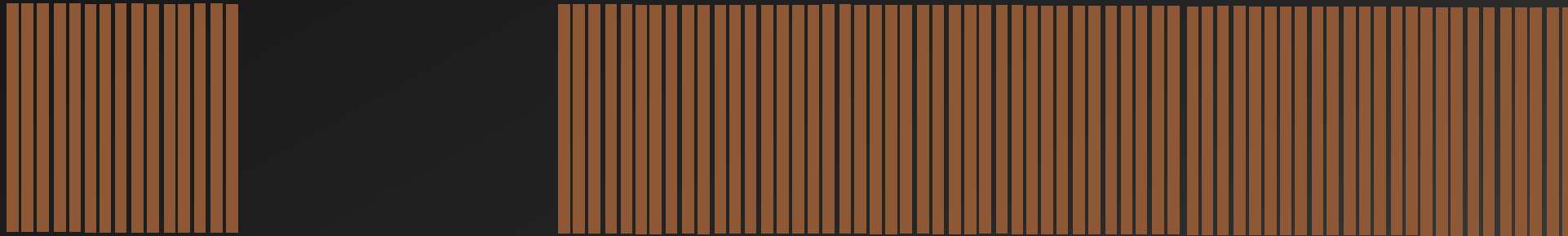
d | Was bedeutet 80% Qualität?

Die Verteilung der Lücken könnte jedoch auch anders sein ...



d | Was bedeutet 80% Qualität?

Noch größer wird die Lücke, wenn die 20% am Stück fehlen ...

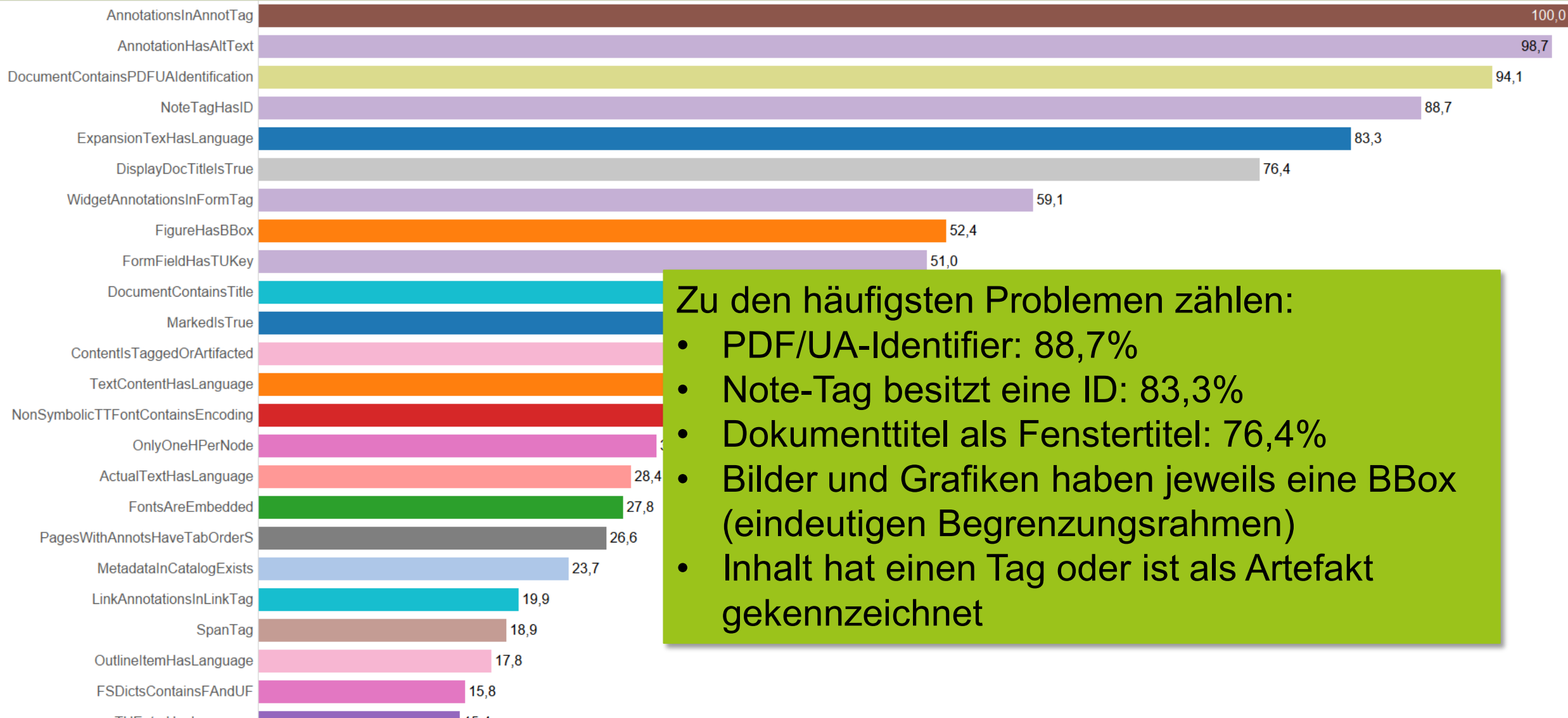


Vertrauen Sie dieser Brücke noch?

4 | Einzelne PDF/UA-Anforderungen

- Rangliste der häufigsten PDF/UA-Probleme
- Beispiel für ein PDF/UA-Problem und dessen Verteilung über die Ministerien hinweg
- PDF/UA-Problempprofile: Typische Fehler für bestimmte Organisationen identifizieren (die auf Prozessfehler hinweisen)

a | Rangliste PDF/UA-Probleme

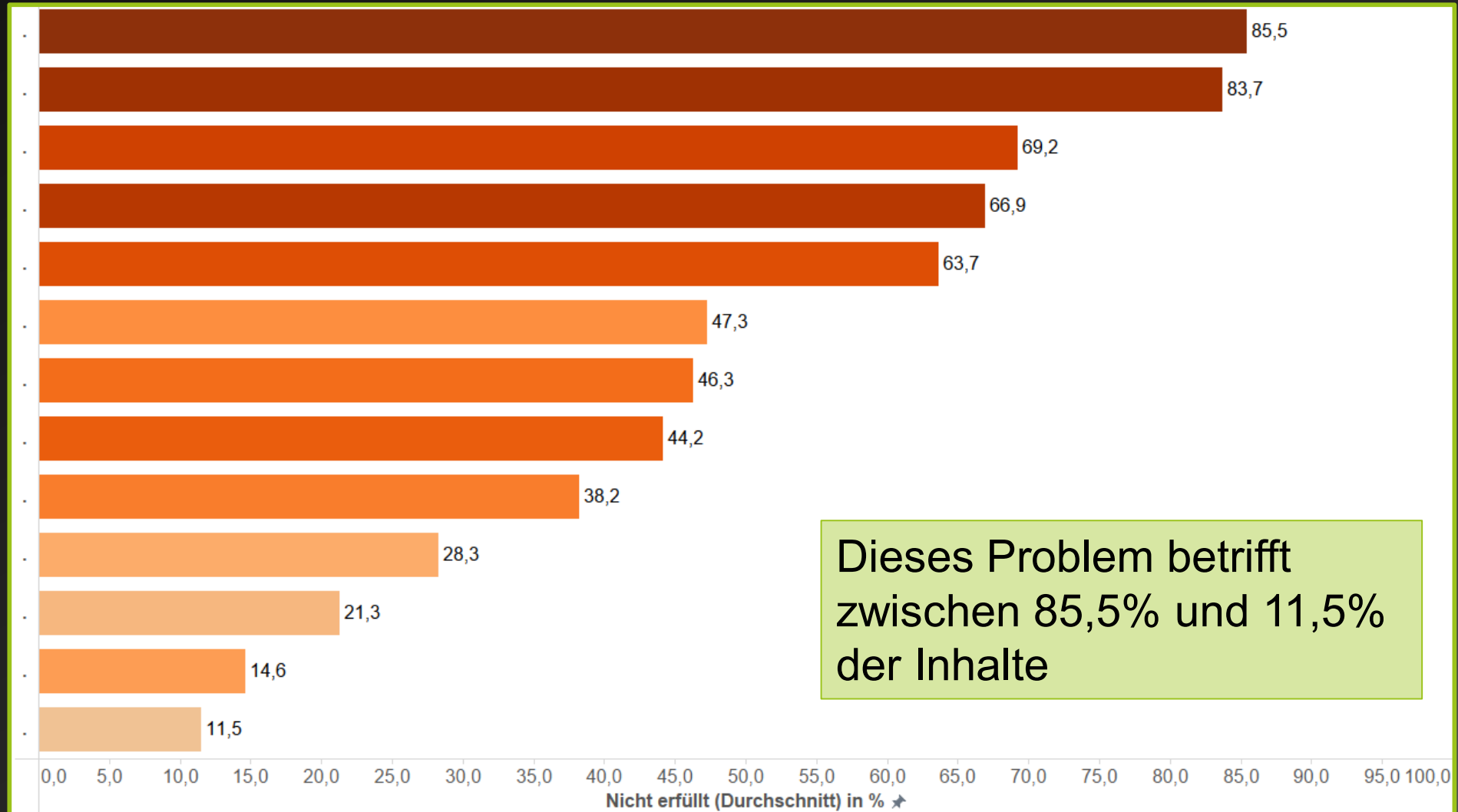


Zu den häufigsten Problemen zählen:

- PDF/UA-Identifizierung: 88,7%
- Note-Tag besitzt eine ID: 83,3%
- Dokumenttitel als Fenstertitel: 76,4%
- Bilder und Grafiken haben jeweils eine BBox (eindeutigen Begrenzungsrahmen)
- Inhalt hat einen Tag oder ist als Artefakt gekennzeichnet

a | Einzelne PDF/UA-Anforderung

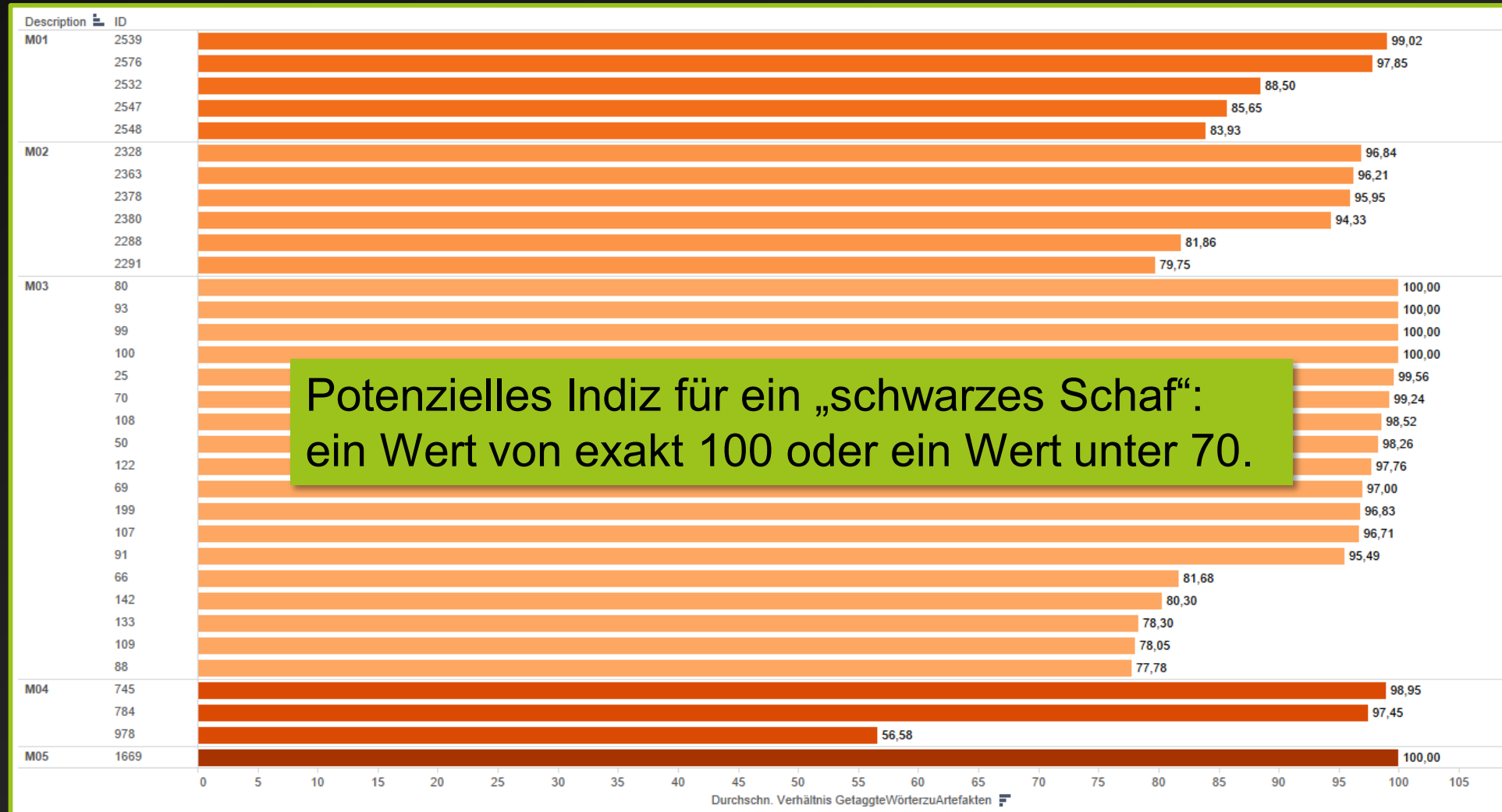
Content is tagged or artifacted | Nicht erfüllt in %



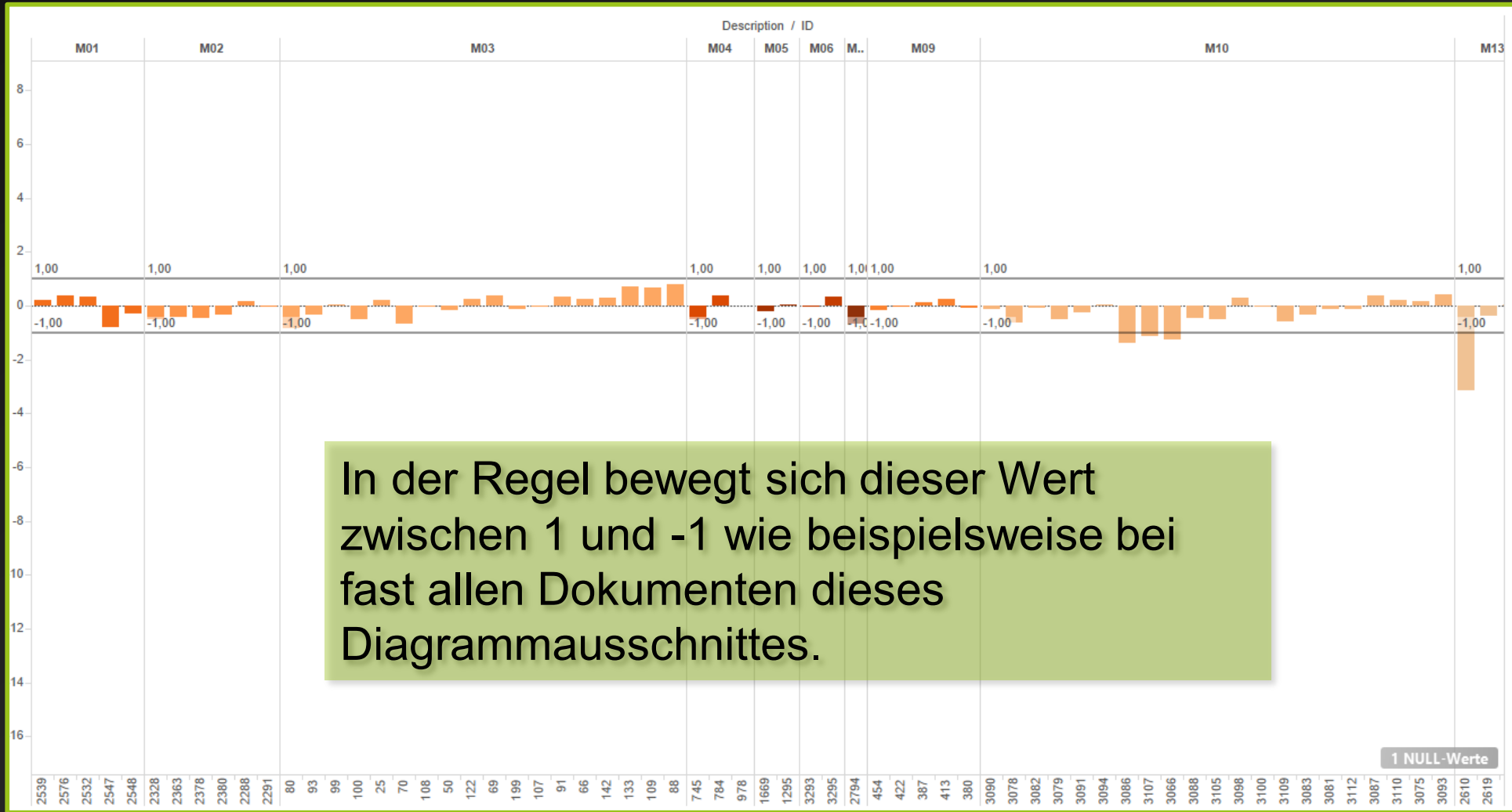
5 | Indizien für “Semantic Richness”

- Anzahl der getaggten Wörter im Vergleich zu Artefakten
- Anzahl Tags (außer Containerelemente und Span) im Vergleich zu Formatierungsänderungen

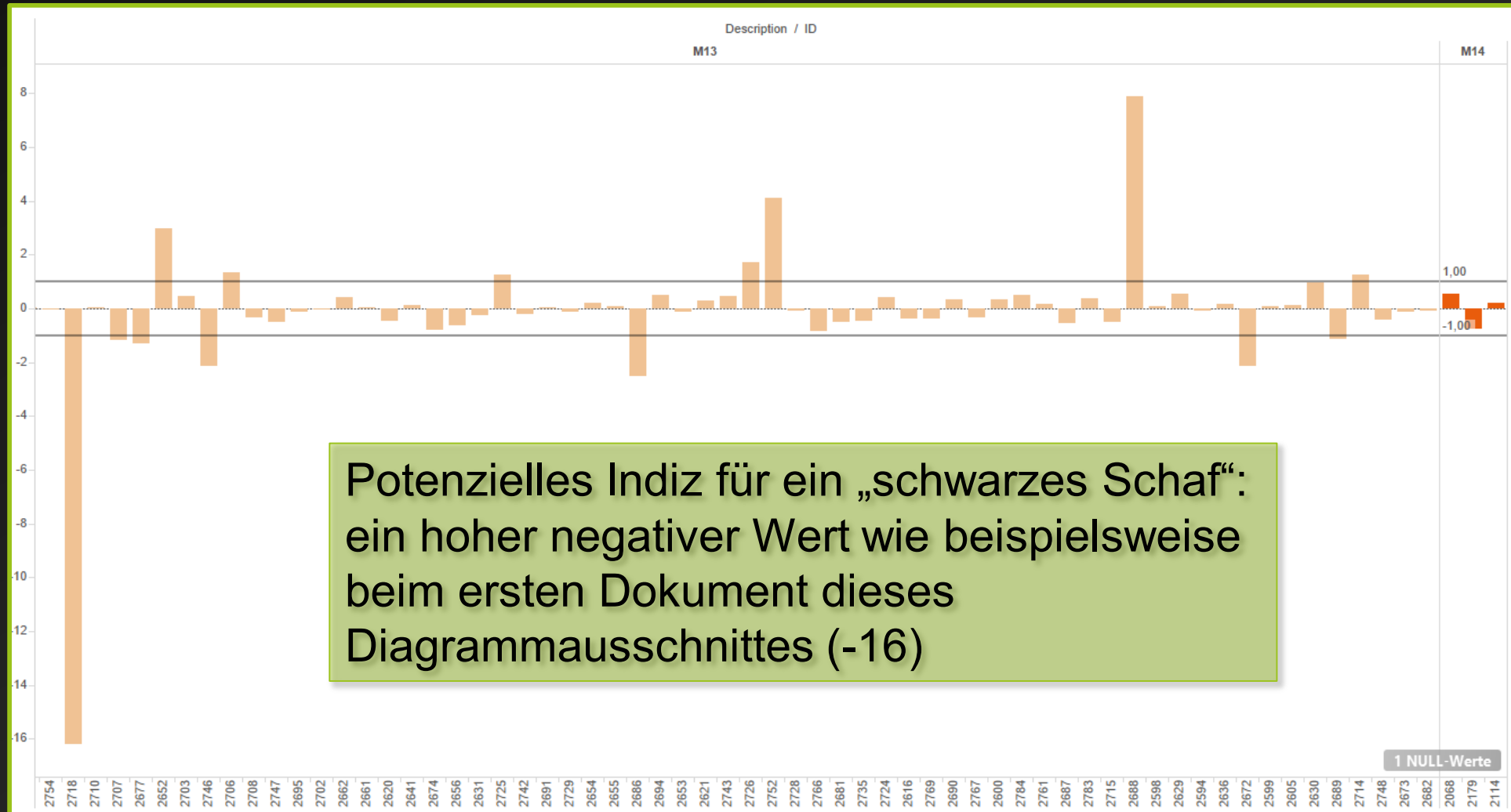
a | Anteil getaggter Wörter vs Artefakte in %



b | Anzahl Tags vs Formatierungsänderungen (1)



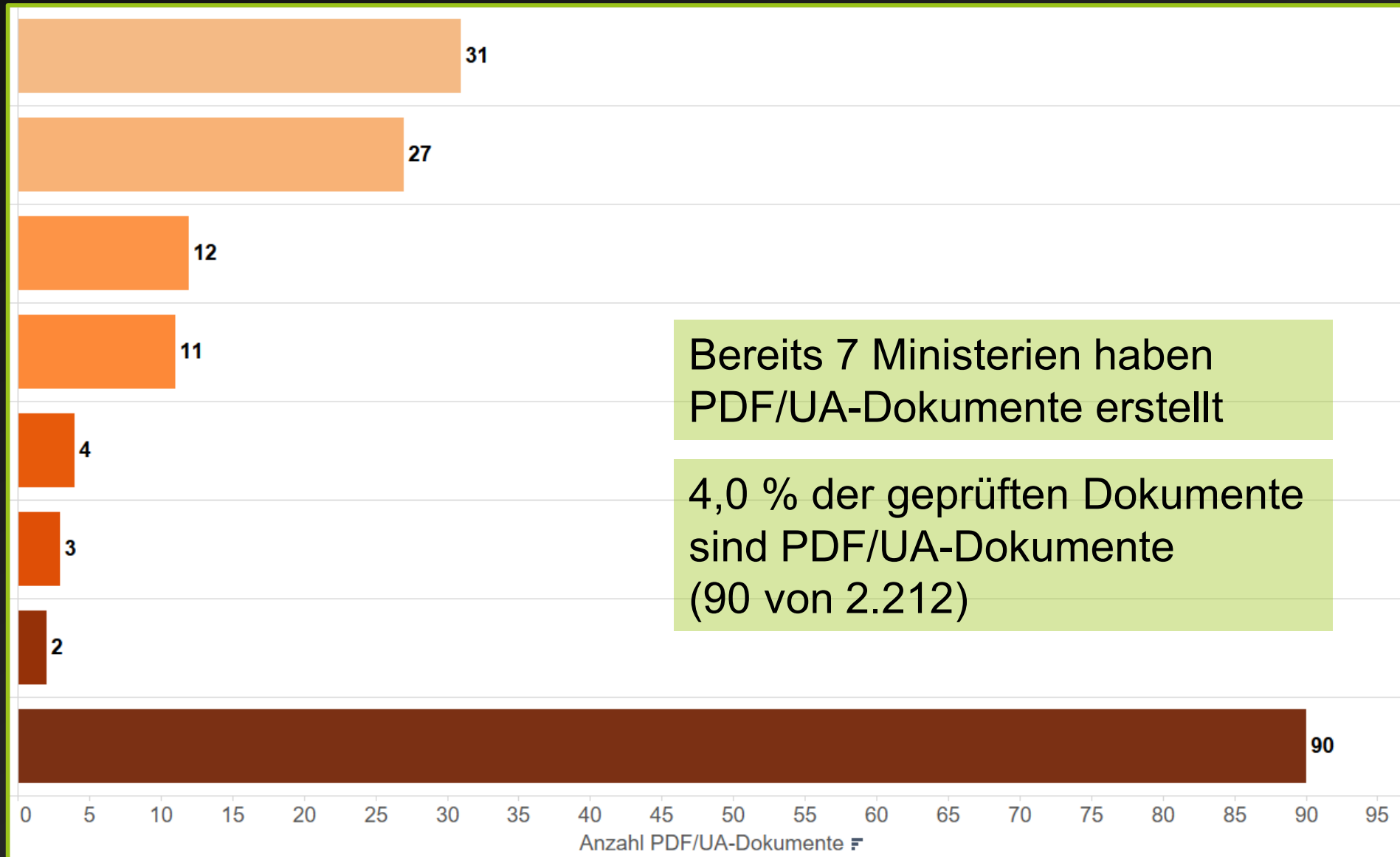
b | Anzahl Tags vs Formatierungsänderungen (1)



6 | 100% PDF/UA-Dokumente

- Wieviele Dokumente erfüllen bereits alle Maschinen-prüfbare PDF/UA-Anforderungen?

a | 100% PDF/UA-Dokumente (Anzahl)





III. Fazit und Ausblick

1. Zusammenfassung der wichtigsten Ergebnisse und Ausgangsfragen aufgreifen
2. Ausblick: Welche neuen Möglichkeiten eröffnen sich?
3. Interesse an weiteren Ergebnissen?

1 | Zusammenfassung der Ergebnisse (1)

Stand der PDF-Barrierefreiheit

- 56,1% der 2.212 analysierten Dokumente sind PDF mit Tags
- Sehr große Unterschiede zwischen den Ministerien
- Durchschnittliche Maschinen-prüfbare PDF/UA-Qualität insgesamt unter 80%
- Ziel wäre: >95%

Ist PDF/UA angekommen? Ist PDF/UA erreichbar?

- Ja, es werden bereits 0-Fehler-Dokumente umgesetzt

1 | Zusammenfassung der Ergebnisse (2)

Die häufigsten PDF/UA-Probleme

- Nicht vollständig getagged, fehlende Auszeichnung als Artefakte, Fehlender PDF/UA-Identifizierer, Fehlende BBox, ... weisen auf mangelhafte Werkzeuge hin. Ließe sich relativ einfach beheben.

Effizientes Monitoring & „Schwarze Schafe“ schnell identifizieren

- Typische Fehler und besonders fehlerhafte Dokumente schnell erkennbar

Fortschritte messbar machen & Basis für Vergleiche

- Monitoring-Technologie bietet Möglichkeit von regelmäßigen Erhebungen, um den Stand der PDF-Barrierefreiheit und eventuell erzielte Fortschritte zu quantifizieren und zu visualisieren

2 | Ausblick: Außenperspektive

- PDF/UA sichtbar machen
- Diejenigen Organisationen sichtbar machen, die den DIN-/ISO-Standard bereits erfüllen
- Regelmäßiger PDF/UA Real Life Check für ein Ranking von Institutionen, Organisationen und Ländern
- **Übrigens:**
Die nächsten Studien werden nicht mehr anonymisiert sein

2 | Ausblick: Innenperspektive

- Effizientes internes PDF/UA-Monitoring
(als wesentlicher Baustein der Qualitätssicherung)
- Typische PDF/UA-Fehler über eine beliebige Dokumentmenge lassen sich schnell identifizieren
- Ansatzpunkte für Optimierungen von Prozessen und Werkzeugen
- Kombinierbar mit Lösungen zum Dokumentenmanagement

3 | Interesse an weiteren Ergebnissen?

- Ministerien können Einblick in ihre Detailergebnisse erhalten
- Publikation ist in Vorbereitung

Fragen?



Sprechen Sie uns an!