Kibana als Werkzeug zur Unterstützung der Metadatenkuration

Kolloquium Wissensinfrastruktur, 26.1.2018

AGENDA

1) Motivation

   Metadaten in OpenAIRE

2) Kibana

   Indexierung, Suche & Analyse, Schwierigkeiten

3) Referenzen

Motivation / Projekt OpenAIRE2020

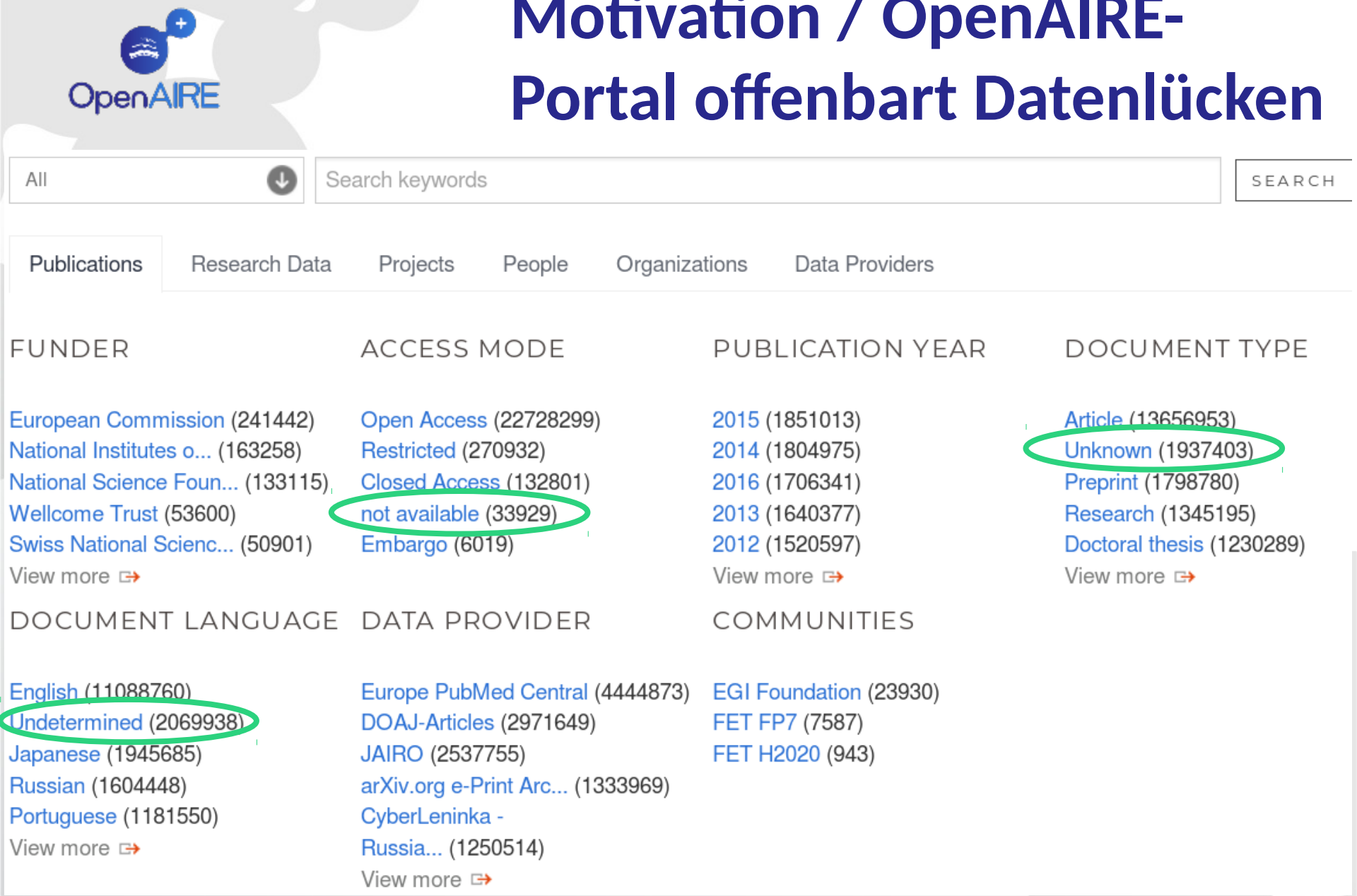
- promote open scholarship and improve discoverability and reusability of research publications
- OpenAIRE platform / technical infrastructure interconnecting collections of research outputs across Europe
- create workflows and services on top of this repository content
- assist in monitoring [EC funded] research outputs, collaboration with national funders

[https://www.openaire.eu/project-factsheets]

Motivation / OpenAIRE aggregiert heterogene Metadaten

Plattformen:	**OJS, DSpace, EPrints, …**
Länder:	**Europa, Asien, Amerika, Afrika, Australien**
Quellen:	**Aggregator, Institut, Journal, …**
Publikationen:	**Literatur, Forschungsdaten, Software, …**
Sprachen:	**viele! betrifft Publikationen und Metadaten**
Formate:	**Dublin Core, DateCite, JATS, …**

Motivation / OpenAIRE-Portal offenbart Datenlücken



Motivation / Datenanalyse fördert Datenqualität

Analyse nativer Daten (nicht transformierter) hilft:

- **Fehler zu entdecken**
- **fruchtbare Inhalte zu entdecken**
- **Formate zu vergleichen**
- **…**

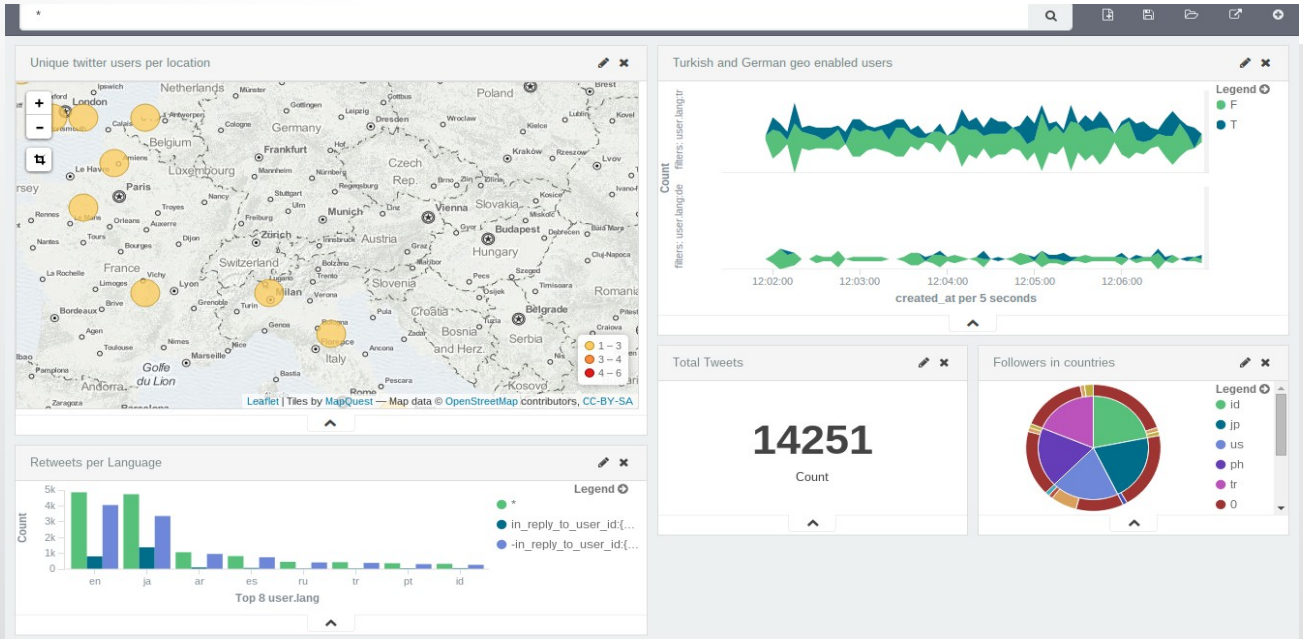
damit lässt sich:

- **die OpenAIRE-Transformation verbessern**
- **den Datenquellen Rückmeldung geben**

Kibana / das OpenSource-Tool hilft bei Metadatenanalyse

Kibana erlaubt, Metadaten
~ zu inspizieren & suchen
~ zu evaluieren & visualisieren

Kibana basiert auf
~ Elasticsearch
~ JSON

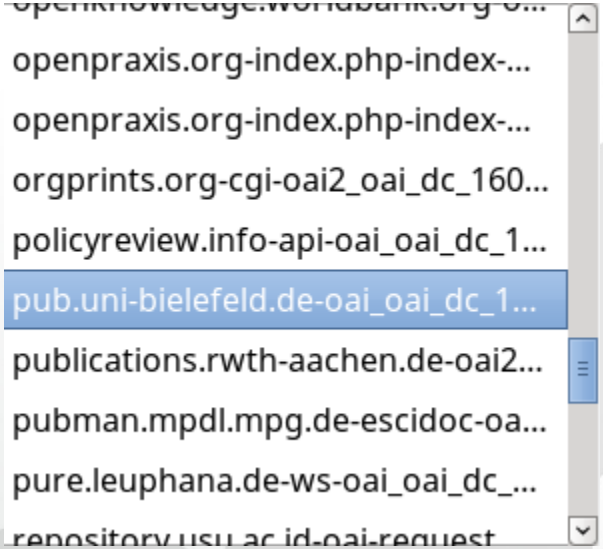


Kibana Import / Indexierung von Metadaten via JSON

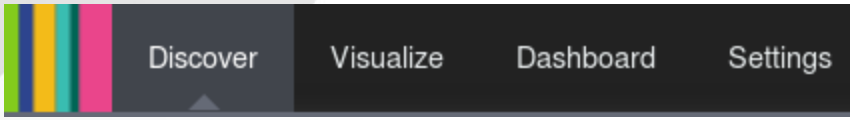
zum Import von Metadaten in Kibana werden

- **Datenquellen 'geharvested'**
- **die Metadaten nach JSON transformiert (XML.toJSONObject)**
- **die Metadaten in Elasticsearch indexiert**

Kibana kann viele Indizes enthalten, und individuell oder zusammen analysieren



Kibana Modi / Kibana bietet 4 Sichten



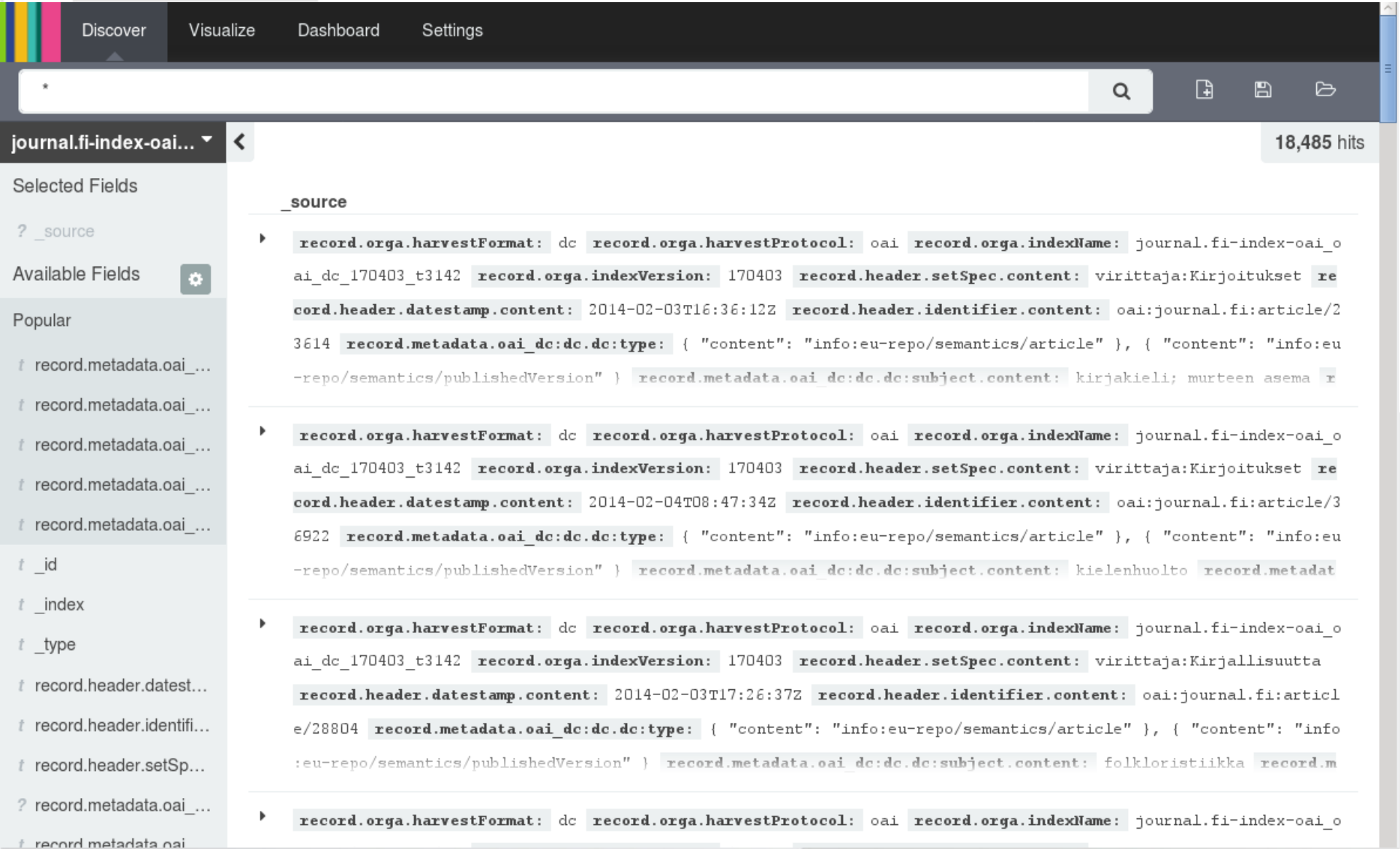
Discover: zeigt Datensätze als Liste/Tabelle

Visualize: visualisiert Daten mit Graphen, Karten, Tabellen, ...

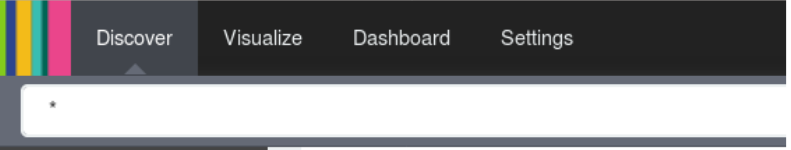
Dashboard: stellt Visualisierungen bereit

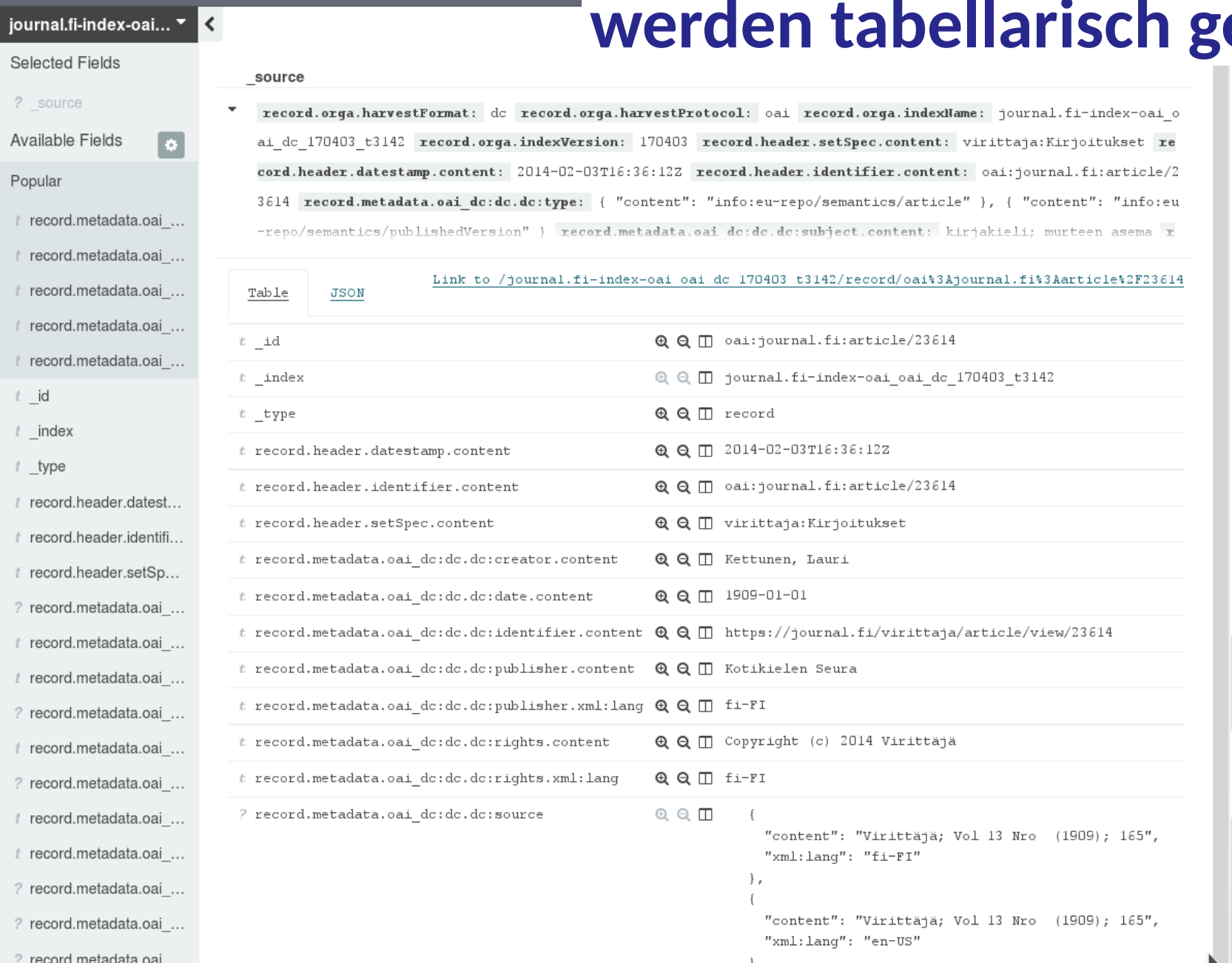
Settings: fügt Indizes hinzu, zeigt Indexierungsdetails

Kibana Discover / Discover-Sicht durchsucht Datensätze

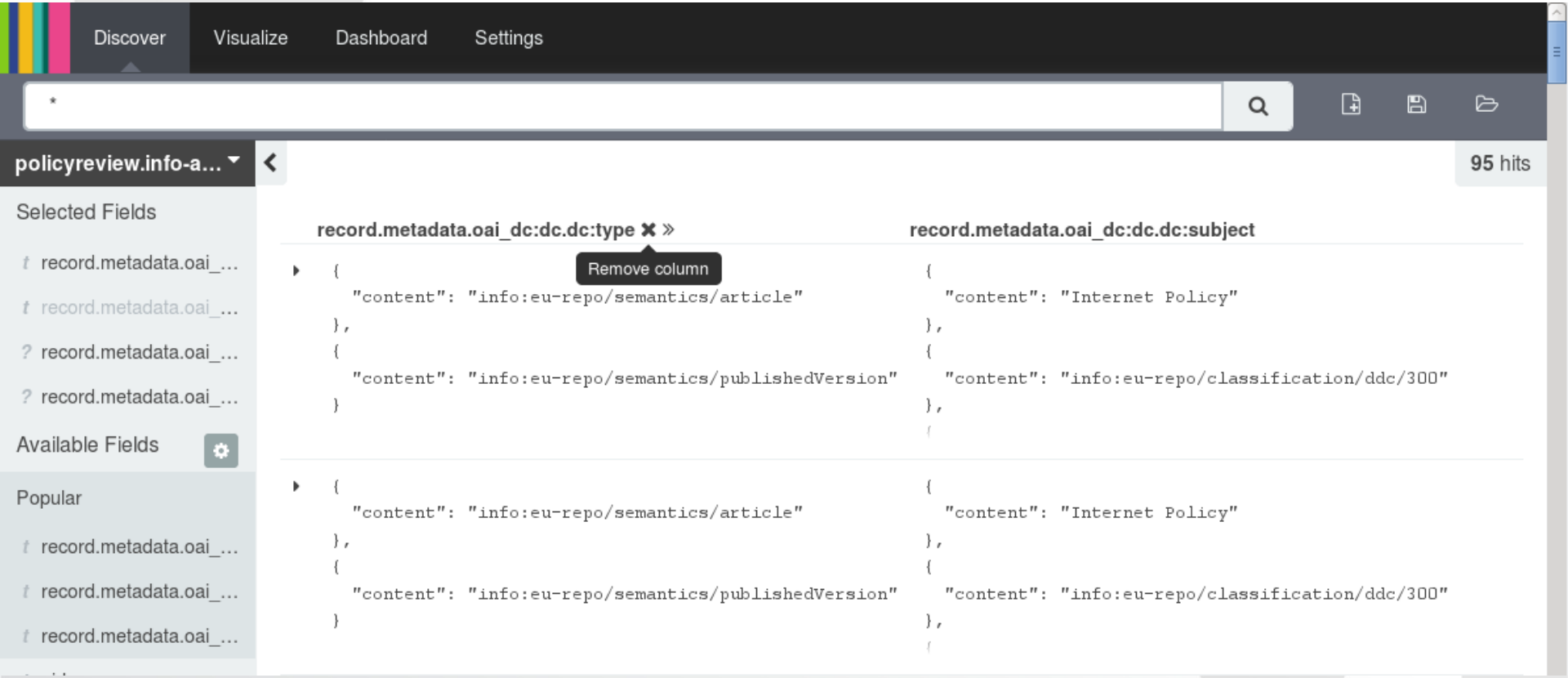


Kibana Discover / Records werden tabellarisch gezeigt





Kibana Discover / Felder lassen sich im Überblick betrachten

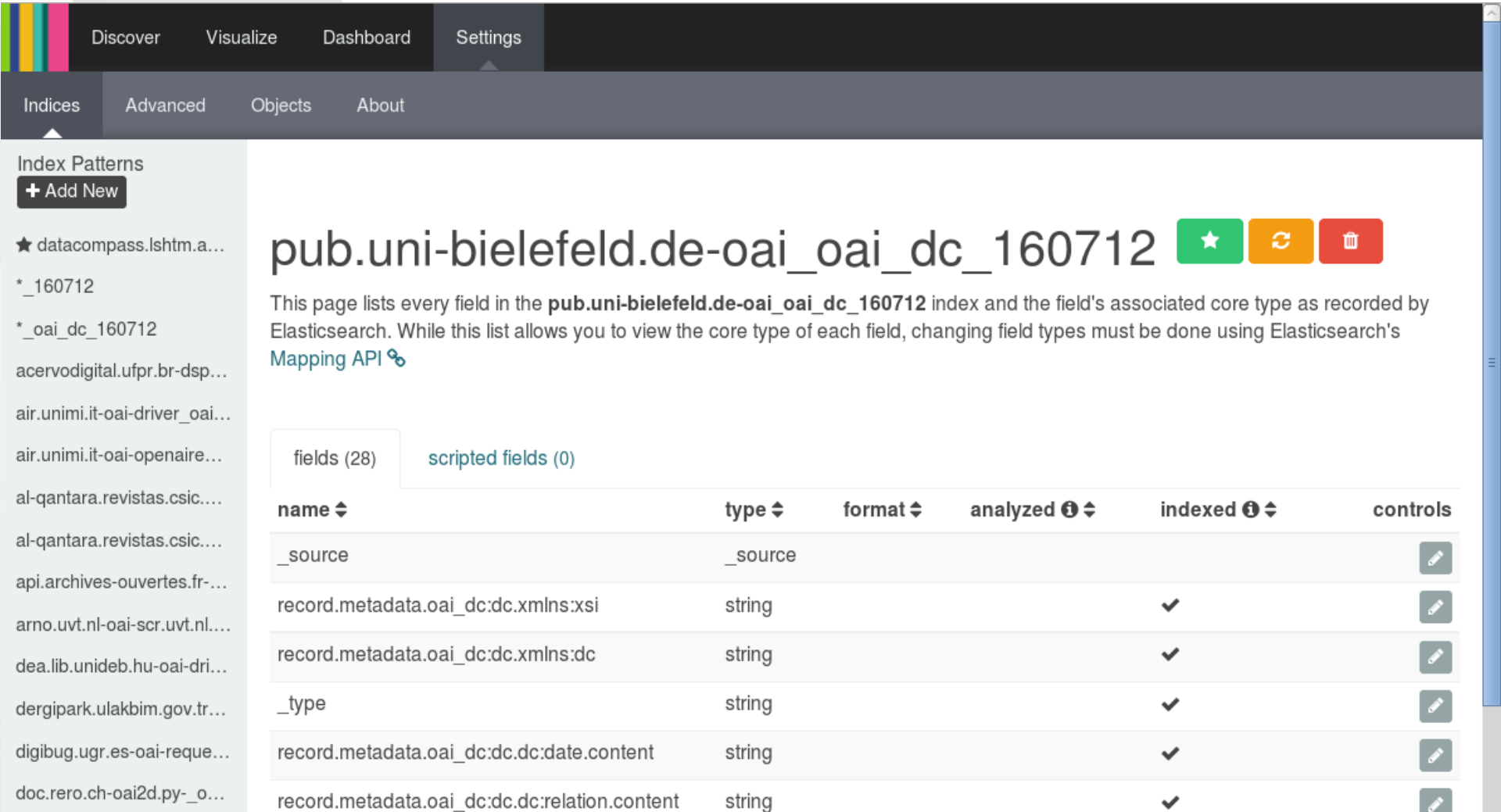


Kibana Discover / gesucht wird nicht in orig. Feldwerten sondern im Index

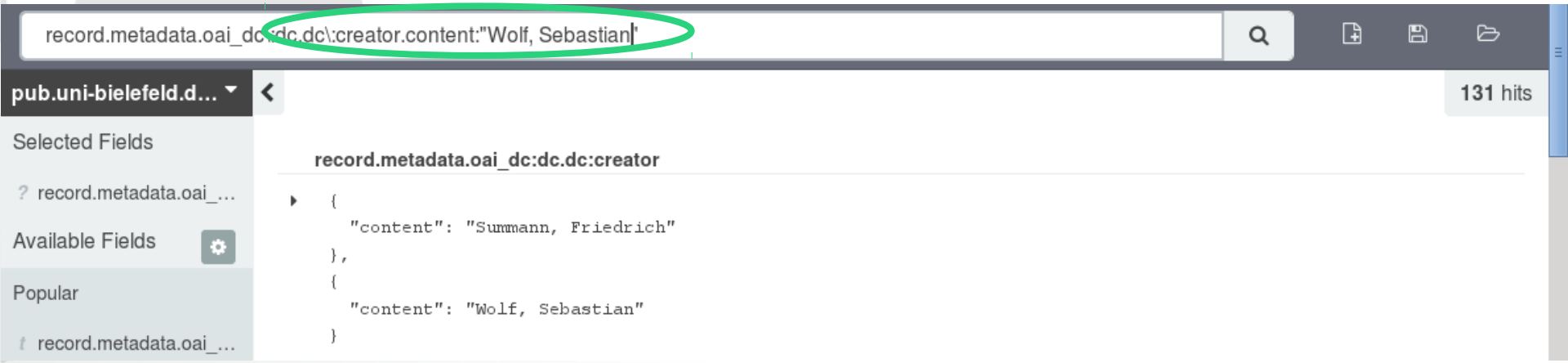
Suche wird beeinflusst durch

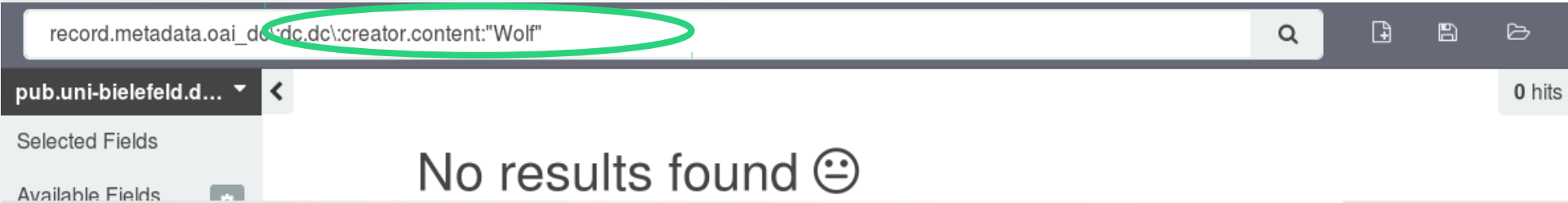
- **Typisierung der Feldwerte (interpretiert als String, Zahl, Datum, …)**
- **eventuelle Analyse der Feldwerte (zerlegt bei Satz-/Leerzeichen, Kleinschreibung)**

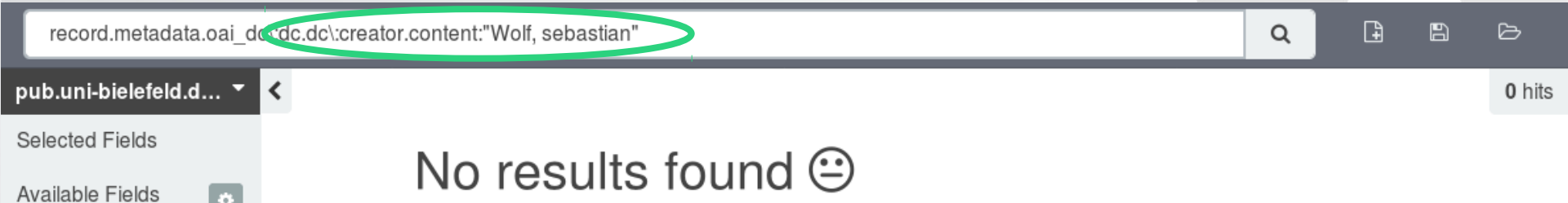
Kibana Settings / Settings-Sicht zeigt je Index Indexierungsdetails der Felder



Kibana Discover / in nicht-analysierten Feldern muß Suche genau sein







Discover / gesuchte Werte lassen sich verschieden beschreiben

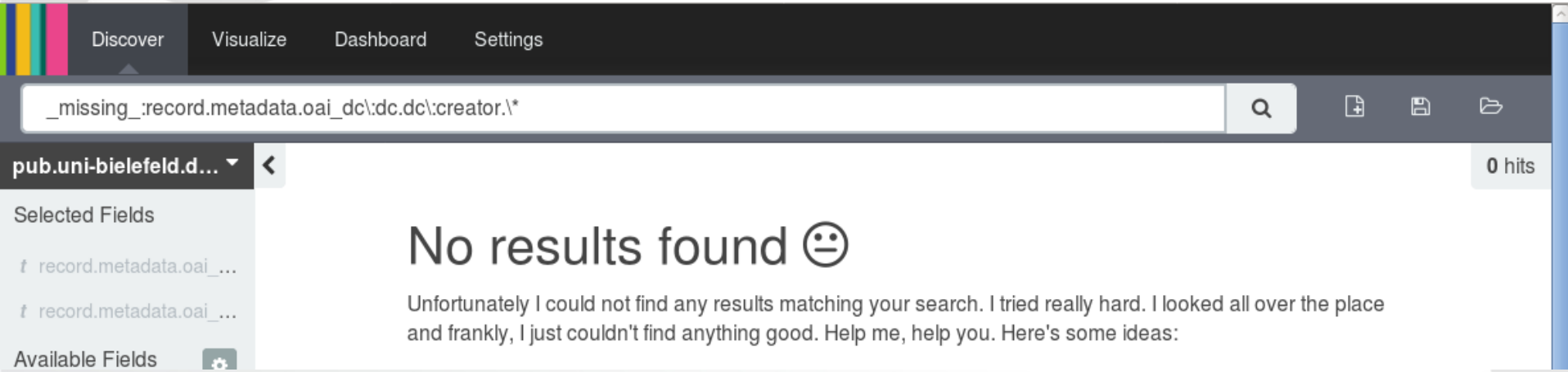
Suchoptionen:

Term:	**hun**	**(ungarisch)**
Phrase:	**"..."**	
Wildcard:	**urn\:issn\:***	
regulärer Ausdruck:	**/(ger|ita)/**	**(deutsch oder italienisch)**

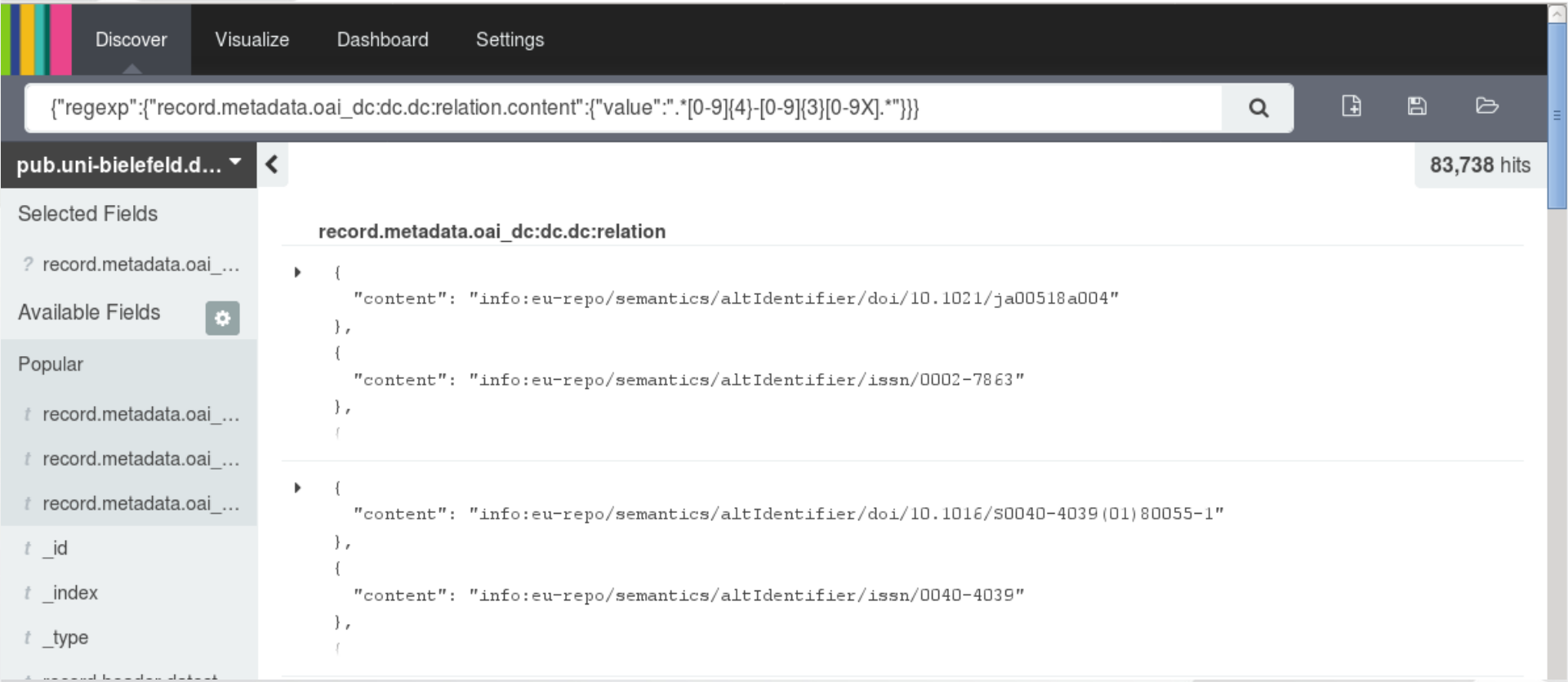
Kibana Discover / Suchen lassen sich auch komplexer gestalten

bestimmtes Feld:	**"dc:title":..., dc\:title:...**
fehlende Felder:	**_missing_:(dc\:creator OR dc\:contributor)**
existierende Felder:	**_exists_:"dc:creator" or _exists_:"dc:contributor"**
logischer Ausdruck:	**AND, OR, NOT**
JSON:	**{"wildcard":{"record.header.setSpec":"j?urnal:?"}}**

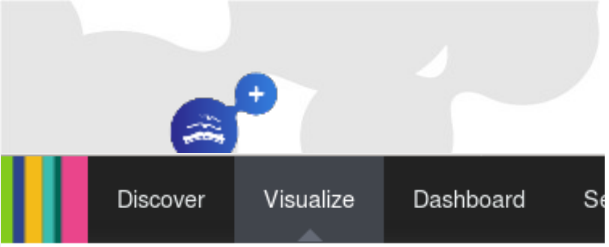
Kibana Discover / fehlen wichtige Felder bzw. Feldwerte?



Kibana Discover / sind z.B. ISSNs gegeben?



Kibana Visualize / Kibana bietet diverse Diagramme/Graphen

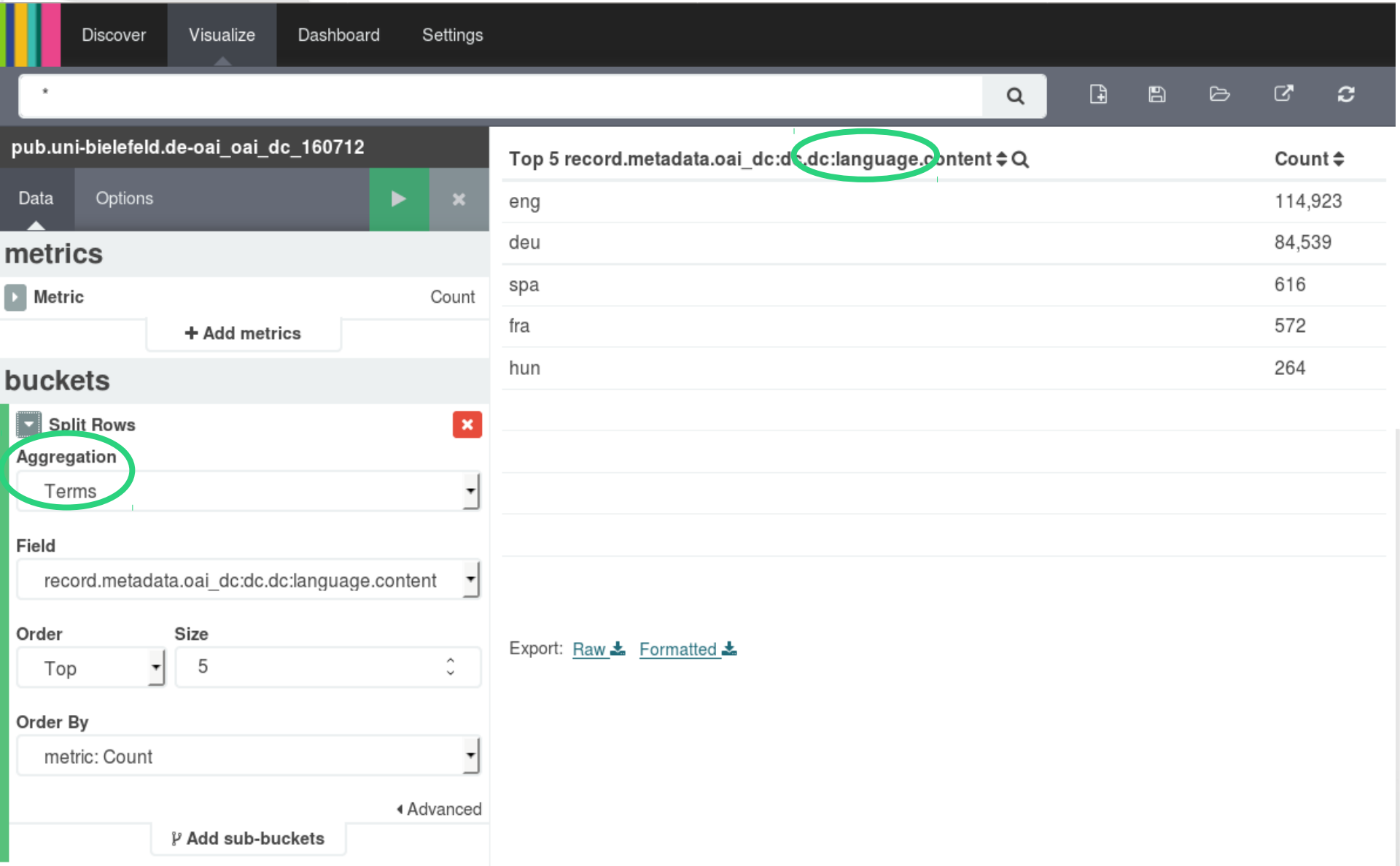


Create a new visualization

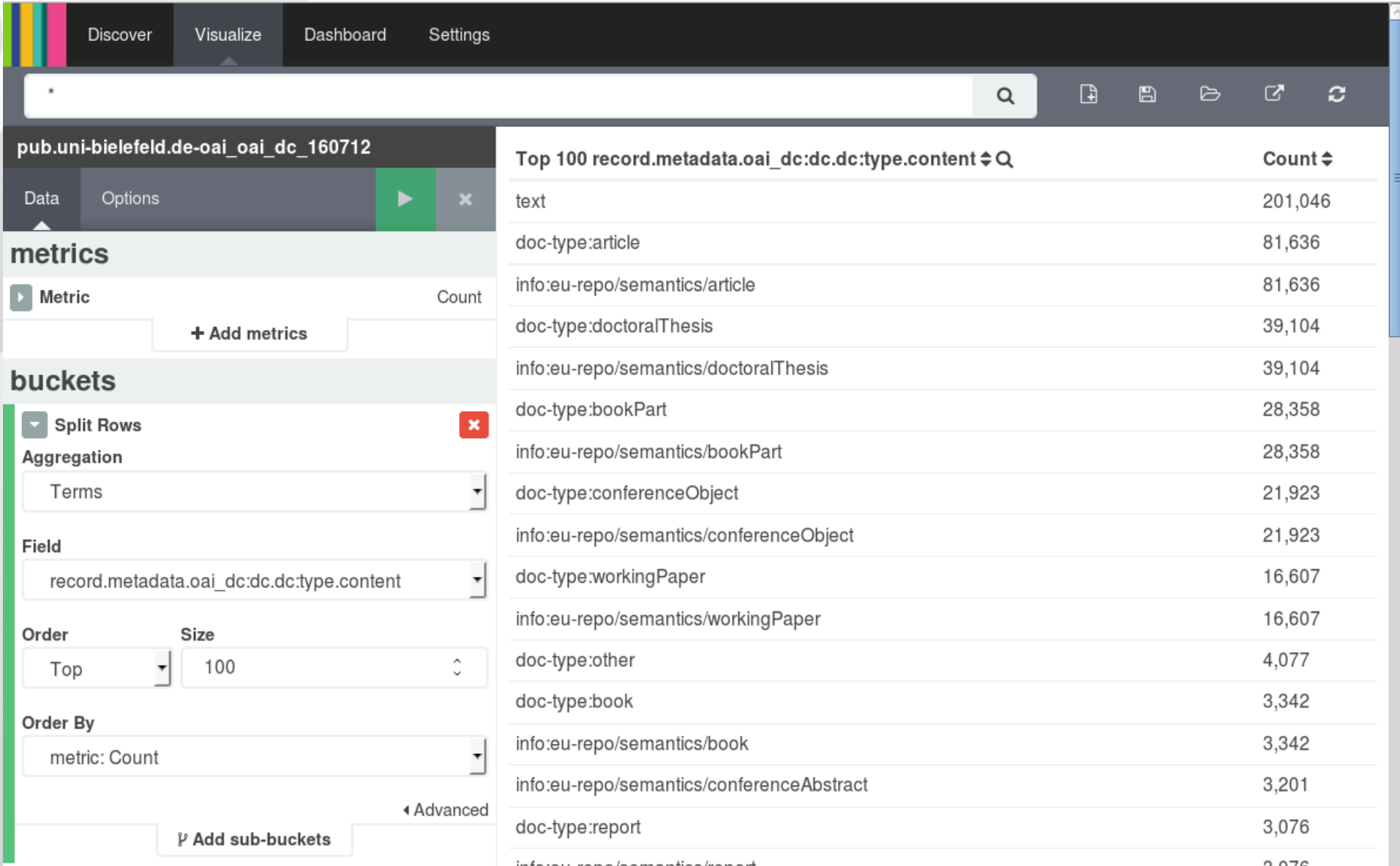
Step 1

Area chart	Great for stacked timelines in which the total of all series is more important than comparing any two or more series. Less useful for assessing the relative change of unrelated data points as changes in a series lower down the stack will have a difficult to gauge effect on the series above it.
Data table	The data table provides a detailed breakdown, in tabular format, of the results of a composed aggregation. Tip, a data table is available from many other charts by clicking grey bar at the bottom of the chart.
Line chart	Often the best chart for high density time series. Great for comparing one series to another. Be careful with sparse sets as the connection between points can be misleading.
Markdown widget	Useful for displaying explanations or instructions for dashboards.
Metric	One big number for all of your one big number needs. Perfect for show a count of hits, or the exact average a numeric field.
Pie chart	Pie charts are ideal for displaying the parts of some whole. For example, sales percentages by department.Pro Tip: Pie charts are best used sparingly, and with no more than 7 slices per pie.
Tile map	Your source for geographic maps. Requires an elasticsearch geo_point field. More specifically, a field that is mapped as type:geo_point with latitude and longitude coordinates.
Vertical bar chart	The goto chart for oh-so-many needs. Great for time and non-time data. Stacked or grouped, exact numbers or percentages. If you are not sure which chart your need, you could do worse than to start here.

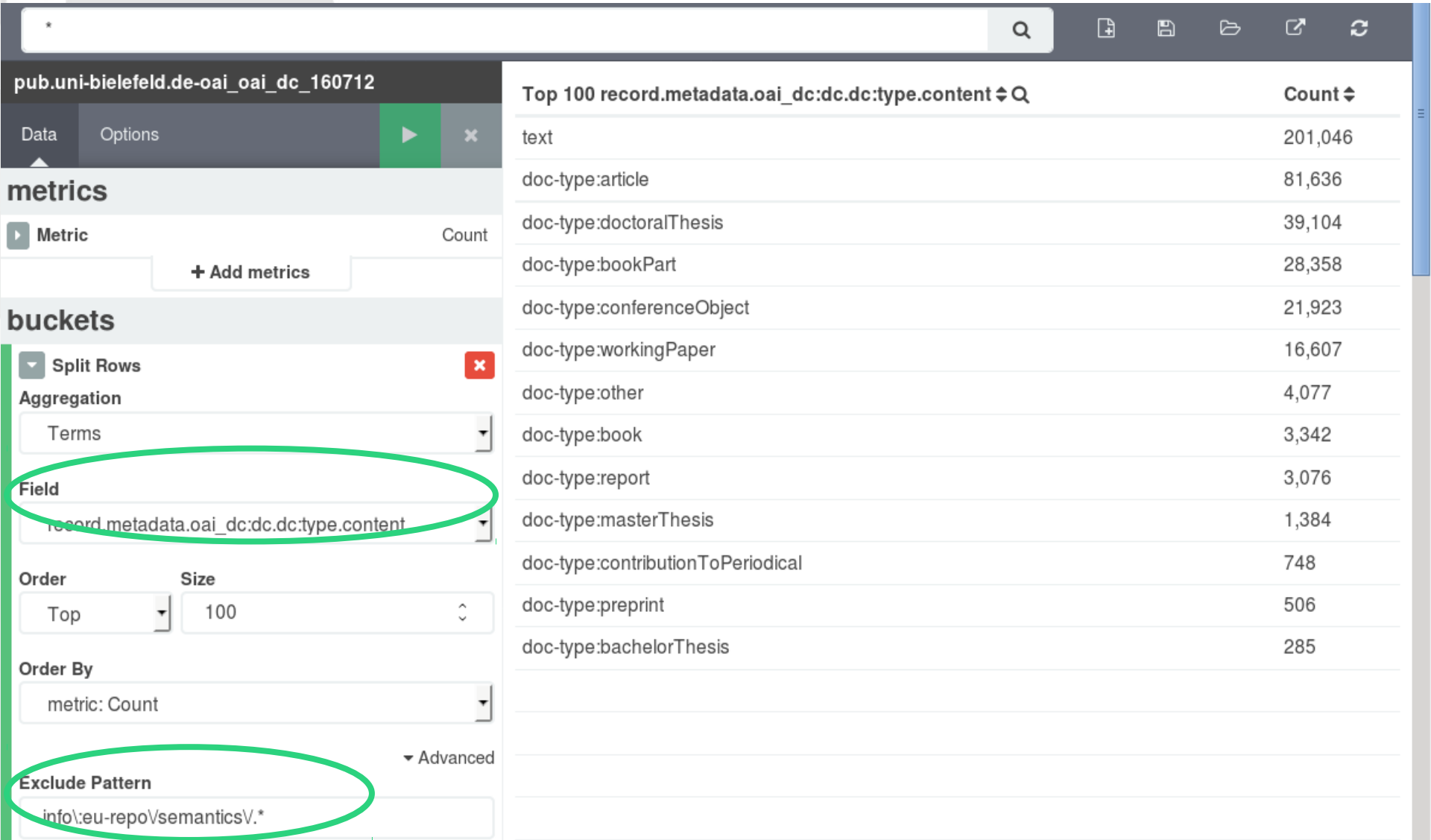
Kibana Visualize / Bucket Aggregation zeigt Feldwerte im Überblick



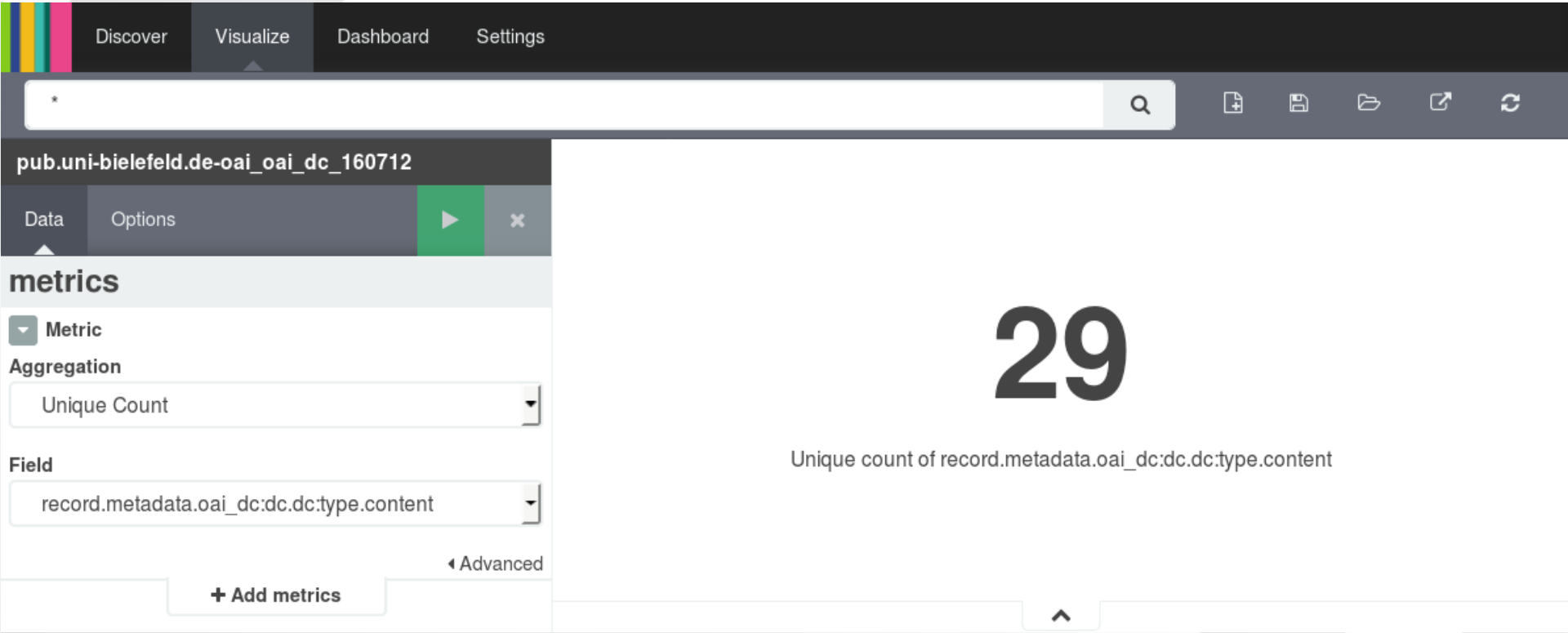
Kibana Visualize / welche Publikationstypen kommen vor?



Kibana Visualize / welche Publikations-typen weichen vom Vokabular ab?



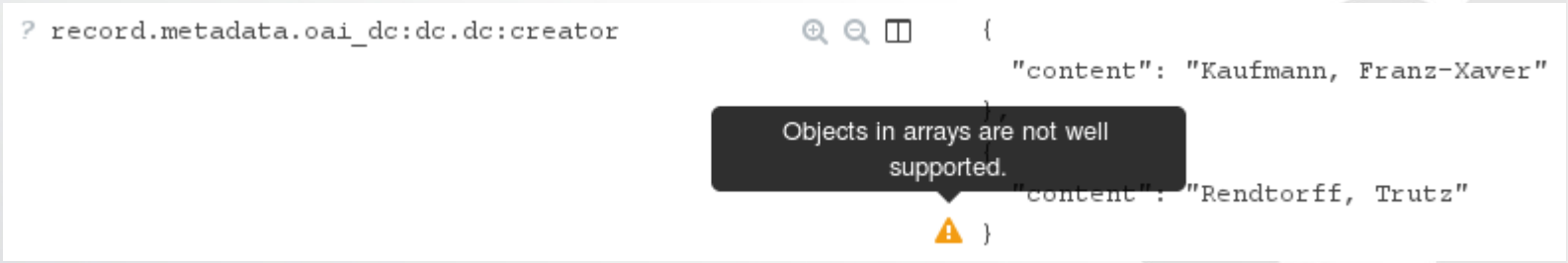
Kibana Visualize / wieviele Publikationstypen kommen vor?



Fazit / zur Analyse von Metadaten ist Kibana sehr nützlich!

gewöhnungsbedürftig ist (noch):

- **Indexierung komplex, via JSON** **- selten nötig**
- **Komplexität erschwert Bedienung** **- erlaubt aber auch viel**
- **GUI reagiert zuweilen etwas störrisch**
- **reguläre Ausdrücke undurchsichtig** **- JSON-Queries nutzen**
- **mehrfache Felder schlechter abfragbar** **- Indexierung anpassen**



Referenzen

Kibana

- **Produkt: https://www.elastic.co/products/kibana**
- **Tutorial: https://www.elastic.co/guide/en/kibana/current/getting-started.html**
- **Tutorial: https://www.timroes.de/2015/02/07/kibana-4-tutorial-part-1-introduction/**

OpenAIRE

- **Portal: https://www.openaire.eu**
- **Suche: https://www.openaire.eu/search/find?keyword=**