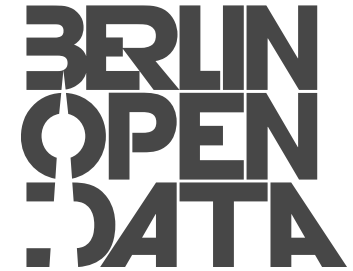


Linked Data mit GitHub

Experimente mit Versionierung, Templates, Actions, Pages zur Veröffentlichung von Linked Data mit GitHub

Hintergrund



Knud Möller

- Linked Data seit 2004 (Uni, Berater, Entwickler)
- Seit 2012: BerlinOnline, verantwortlich für Open-Data-Portal daten.berlin.de

BerlinOnline Stadtportal GmbH & Co. KG

- Technischer Dienstleister des Landes Berlin, insb. für das Hauptstadtportal berlin.de
- Entwicklung, DevOps, Design, Support, Redaktion, Beratung
- Seit 1. Juli 2021 landeseigenes Unternehmen (Tochter des ITDZ)

Linked Data mit GitHub

- Kein offizieller Teil von Berlin Open Data, Experiment/Prototyp, unfertig

Linked Data in Berlin

- **2012** – 1. Berliner Open-Data-Strategie: Handlungsempfehlung zur Einführung von Linked Data als „Datenaufwertung“
- **2018** – Portal-Metadaten als DCAT-AP.de
- **2022** – Beteiligungsprozess zur Erarbeitung einer neuen Open-Data-Strategie für Berlin: Erneut Empfehlung zur Einführung von Linked Data
- **2023** – Ankündigung der Senatsverwaltung für Finanzen: Haushaltsdaten als Linked Data

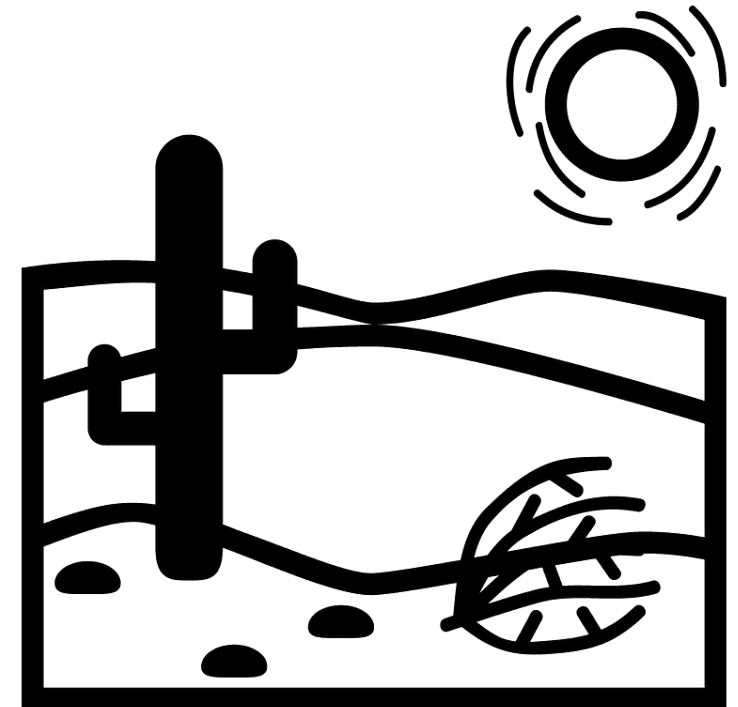
Die ideale Linked-Data-Welt

- Linked Data **direkt von der Quelle.**
- Alle Fachverfahren denken globale **Identifizier, Graph-Strukturen, Vokabulare, Beziehungen zu anderen Daten** etc. von Anfang an mit.
- **Überall RDF!**



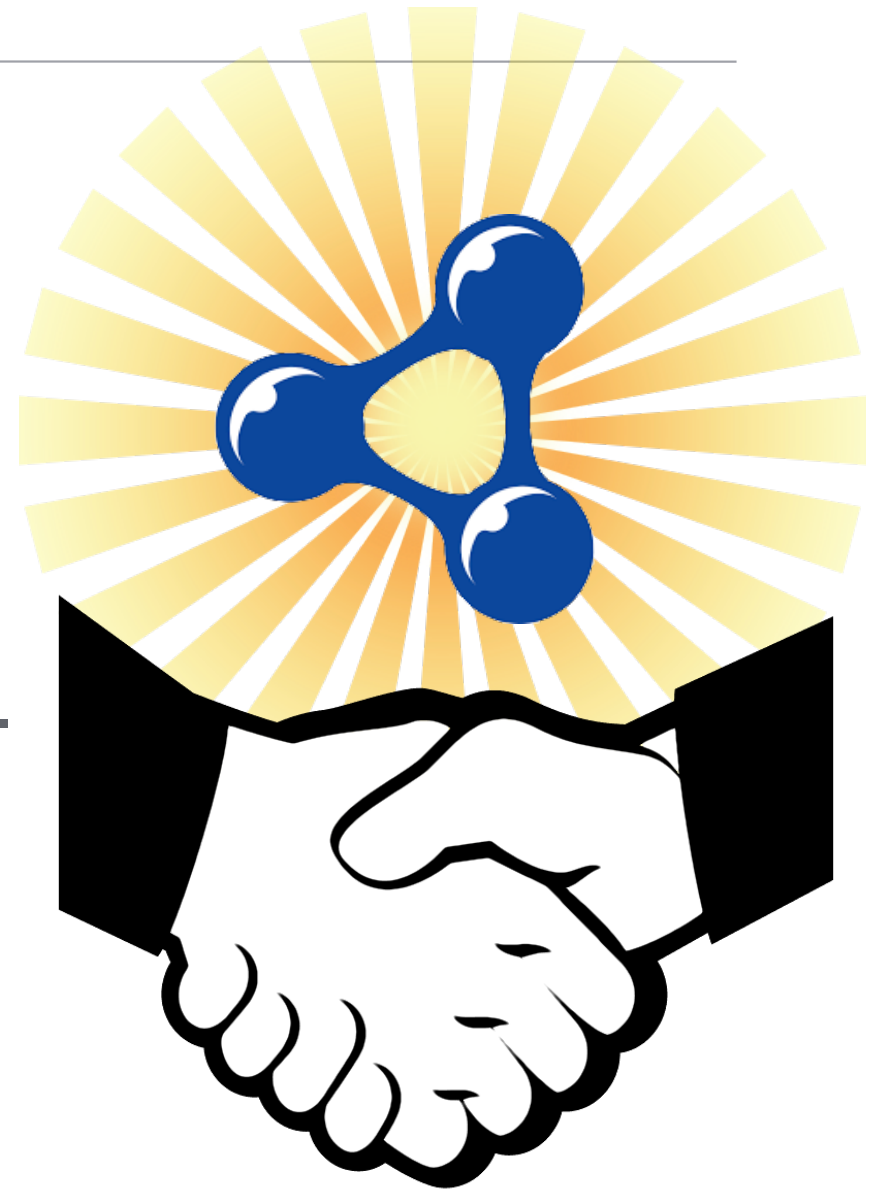
Die Daten-Realität

- Fachverfahren / IT-Systeme teilweise schon **Jahrzehnte im Einsatz**.
- **Relationale Daten, Tabellen**
- **Silo-Denken**
- **Anreiz** für Linked Data für jeden einzelnen Hersteller und Datenbereitsteller **gering**.
- **Linked-Data-Know-How** nicht weit verbreitet.



Kompromiss

- **Nachgelagerte** Linked Data
- Beliebige **Quelldaten aufwerten** („liften“) durch ETL-Prozess (csv2rdf, json2rdf, rdb2rdf etc).
- **Zunächst zentralisiert** (Know-How, Anreiz).
- Beginnen mit zentralen **Referenz-Datensätzen**.
- Später (**gelenkt**) **dezentralisiert** (Standardisierung).



Linked (Open) Data mit GitHub



- GitHub als **Werkzeug** für Linked Data
- (relativ) **niederschwellig**: kostenfrei, keine Ressourcen für Setup und Hosting (Cloud)
- Eigentlich für Code, aber auch gut nutzbar für text-basierte Daten.
- **Version-Control-System**: mit Einschränkung Versionierung von RDF-Daten
- **GitHub-Actions**
- **GitHub-Pages**
- **Repository-Templates**

Linked (Open) Data mit GitHub



- **Minimaler Vendor-Lock-In:**
 - git ist Open Source (Alternativen wie GitLab verfügbar)
 - GH-Actions sind hauptsächlich Docker und Linux
 - GH-Pages ist hauptsächlich Jekyll (OS static site generator)

GitHub Actions



Beliebigen Code bei bestimmten **Triggern** ausführen:

- Ausführung bei GitHub auf Container (Docker)
- Regelmäßige Überprüfung der Quell-Daten (Cron)
- **Transformierung in RDF** bei der Update Quell-Daten
- **Linting, Validierung**
- Erzeugung von **Dokumentation**, statischen **HTML-Seiten** etc.

Referenzdatensatz: Lebensweltlich Orientierte Räume



Die lebensweltlich orientierten Räume (LORs) gliedern das Land Berlin in geografische Einheiten auf vier hierarchischen Stufen, beginnend bei den **12 Bezirken**, über **58 Prognoseräume**, **143 Bezirksregionen**, bis hinunter zu den **542 Planungsräumen**. Die LORs bilden die räumliche Grundlage für Planung, Prognose und Beobachtung demografischer und sozialer Entwicklungen in Berlin.

berlinonline / lod-berlin-lor

<> Code Issues Pull requests Actions Projects Wiki

Berlin's LOR hierarchy as RDF / Linked Open data (01.01.2021)

MIT license

0 stars 0 forks 3 watching Activity

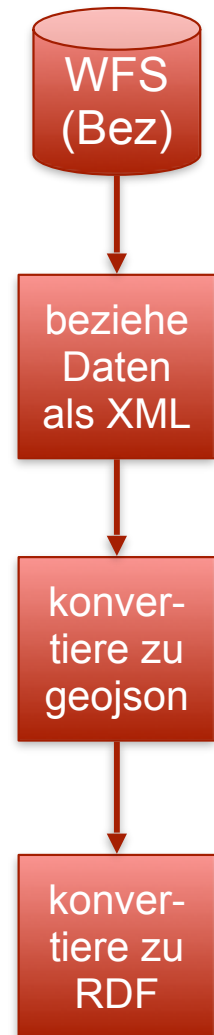
Public repository · Generated from [berlinonline/lod-sg](#)

main

Branches Tags

knudmoeller prevent everything from being built twice ✓ yesterday 29

📁 .github/workflows	yesterday
📁 _data	2 days ago
📁 _includes	2 days ago
📁 _layouts	2 days ago
📁 assets	2 days ago
📁 bin	yesterday
📁 conf	2 days ago
📁 data	2 days ago
📁 query	2 days ago



berlinonline / lod-berlin-lor

<> Code Issues Pull requests Actions Projects Wiki

Berlin's LOR hierarchy as RDF / Linked Open data (01.01.2021)

MIT license

0 stars 0 forks 3 watching Activity

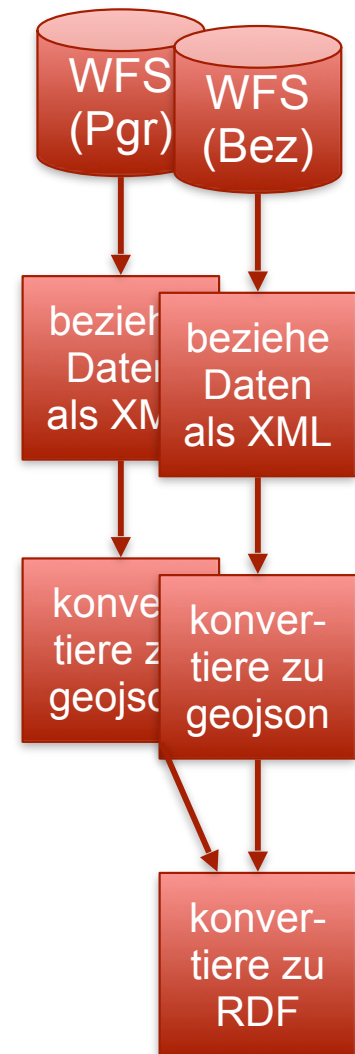
Public repository · Generated from berlinonline/lod-sg

main

Branches Tags

knudmoeller prevent everything from being built twice ✓ yesterday 29

.github/workflows	yesterday
_data	2 days ago
_includes	2 days ago
_layouts	2 days ago
assets	2 days ago
bin	yesterday
conf	2 days ago
data	2 days ago
query	2 days ago



berlinonline / lod-berlin-lor

Code Issues Pull requests Actions Projects Wiki

Berlin's LOR hierarchy as RDF / Linked Open data (01.01.2021)

MIT license

0 stars 0 forks 3 watching Activity

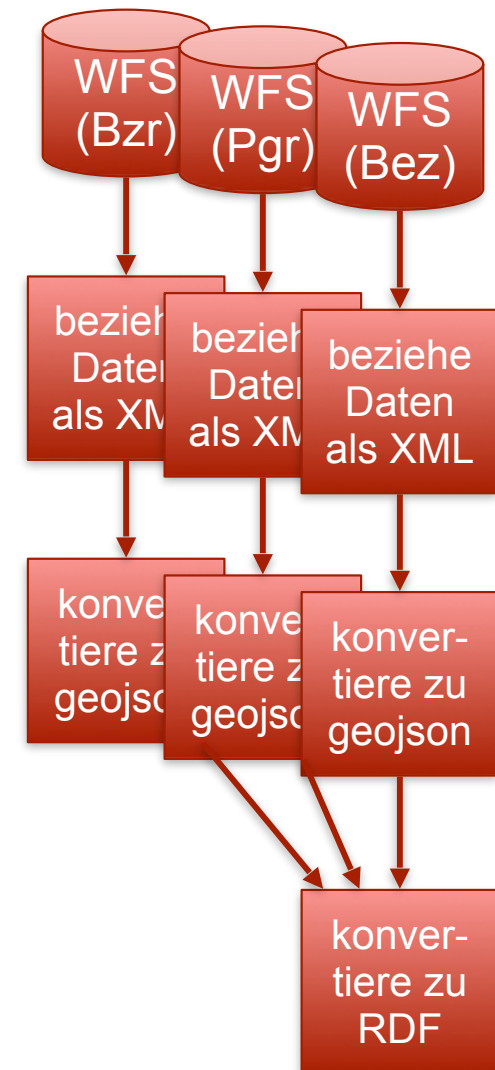
Public repository · Generated from [berlinonline/lod-sg](#)

main

Branches Tags

knudmoeller prevent everything from being built twice ✓ yesterday 29

github/workflows	yesterday
_data	2 days ago
_includes	2 days ago
_layouts	2 days ago
assets	2 days ago
bin	yesterday
conf	2 days ago
data	2 days ago
query	2 days ago



berlinonline / lod-berlin-lor

Code Issues Pull requests Actions Projects Wiki

Berlin's LOR hierarchy as RDF / Linked Open data (01.01.2021)

MIT license

0 stars 0 forks 3 watching Activity

Public repository · Generated from [berlinonline/lod-sg](#)

main

Branches Tags

knudmoeller prevent everything from being built twice ✓ yesterday 29

- .github/workflows yesterday
- _data 2 days ago
- _includes 2 days ago
- _layouts 2 days ago
- assets 2 days ago
- bin yesterday
- conf 2 days ago
- data 2 days ago
- query 2 days ago



berlinonline / lod-berlin-lor

<> Code Issues Pull requests Actions Projects Wiki

Berlin's LOR hierarchy as RDF / Linked Open data (01.01.2021)

MIT license

0 stars 0 forks 3 watching Activity

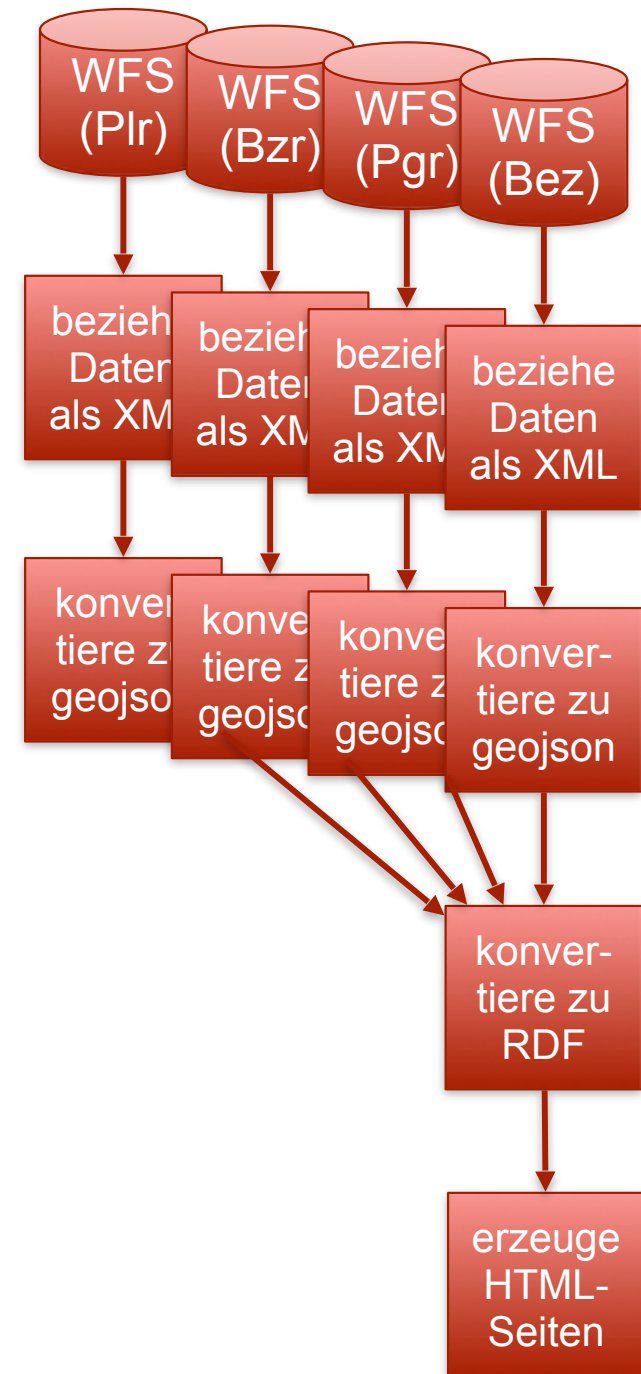
Public repository · Generated from [berlinonline/lod-sg](#)

main

Branches Tags

knudmoeller prevent everything from being built twice ✓ yesterday 29

- .github/workflows yesterday
- _data 2 days ago
- _includes 2 days ago
- _layouts 2 days ago
- assets 2 days ago
- bin yesterday
- conf 2 days ago
- data 2 days ago
- query 2 days ago



GitHub-Pages

- Erzeugen und Hosten einer **statischen Linked-Data-Site** für das Repository
- Dadurch möglich: **Erreichbare (dereferenzierbare) URIs!**
 - <https://www.w3.org/DesignIssues/LinkedData.html>
 - https://data.europa.eu/sites/default/files/d2.1.2_training_module_2.3_persistent_uri_design_and_management_de_edp.pdf
- Basis-URL der GH-Pages ergibt sich aus Basis-URL des Repositories:
 - <https://github.com/berlinonline/lod-berlin-lor/> → <https://berlinonline.github.io/lod-berlin-lor/>
- Basis-URL der GH-Pages ist auch Basis-URL aller RDF-Daten

Lebensweltlich Orientierte Räume (LOR) als RDF (LOR Version 01.01.2021)

Datensatz

Die lebensweltlich orientierten Räume (LORs) gliedern das Land Berlin in geografische Einheiten auf vier hierarchischen Stufen, beginnend bei den 12 Bezirken, über 58 Prognoseräume, 143 Bezirksregionen, bis hinunter zu den 542 Planungsräumen. Die LORs bilden die räumliche Grundlage für Planung, Prognose und Beobachtung demografischer und sozialer Entwicklungen in Berlin. Dieser Datensatz bildet die Daten der bestehenden WFS-Dienste als Linked Data bzw. RDF ab.

Typ:	void:Dataset
zuerst erstellt am:	2022-12-31
zuletzt geändert am:	2023-03-28
Veröffentlichende Stelle:	BerlinOnline Stadtportal GmbH & Co. KG
Homepage:	https://github.com/berlinonline/lod-berlin-lor
Einstiegsressource:	lor:berlin
Beispielressource:	lor:plr_02500729
Beispielressource:	lor:pgr_0250
Beispielressource:	lor:bzr_025007
Beispielressource:	lor:bez_02
Andere Version:	lor19:

Datensatz- <https://berlinonline.github.io/lod-berlin-lor/>

Lebensweltlich Orientierte Räume (LOR) als RDF (LOR V 01.01.2021)

Datensatz

Die lebensweltlich orientierten Räume (LOR) sind geographische Einheiten auf vier hierarchischen Ebenen: Bezirke, über 58 Prognoseräume, 143 Bezirksplanungsgebiete und 542 Planungsräume. Die LORs bilden die Basis für die Prognose und Beobachtung demografischer Entwicklungen in Berlin. Dieser Datensatz bildet die Daten des Linked Data bzw. RDF ab.

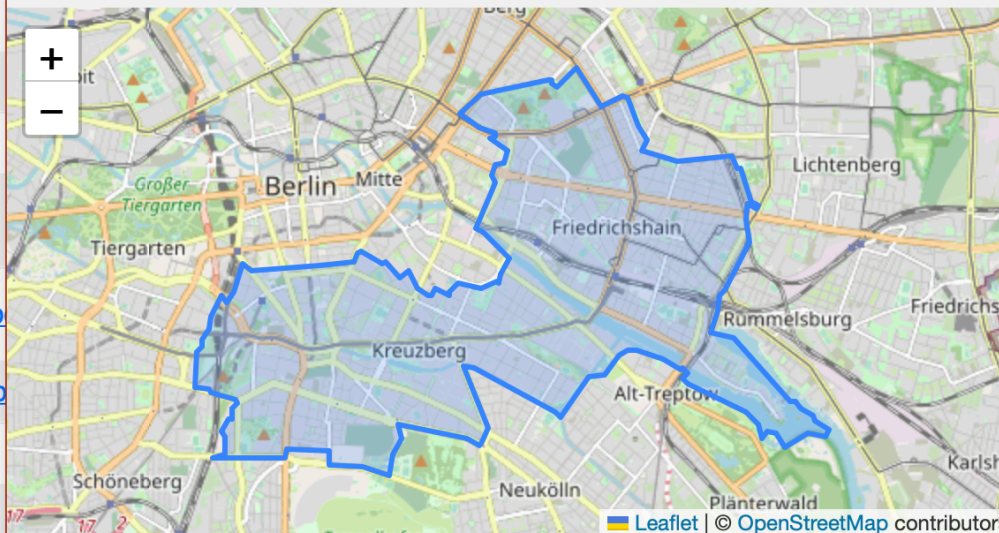
Typ:	void:Dataset
zuerst erstellt am:	2022-12-31
zuletzt geändert am:	2023-03-28
Veröffentlichende Stelle:	BerlinOnline Stadtportal
Homepage:	https://github.com/berlinonline/lod-berlin-lor
Einstiegsressource:	lor:berlin
Beispielressource:	lor:plr_02500729
Beispielressource:	lor:pgr_0250
Beispielressource:	lor:bzr_025007
Beispielressource:	lor:bez_02
Andere Version:	lor19:
Datensatz-URL:	https://berlinonline.github.io/lod-berlin-lor/bez_02

Friedrichshain-Kreuzberg

[Datensatz](#) / [Berlin](#) / Friedrichshain-Kreuzberg (Bezirk 02)

Untereinheiten (Prognoseraum)

- [Kreuzberg Nord](#)
- [Kreuzberg Süd](#)
- [Kreuzberg Ost](#)
- [Friedrichshain West](#)
- [Friedrichshain Ost](#)



Andere Version: [lor19:bez_02](#)

[Toggle Local Graph](#)

[Impressum](#)

Lebensweltlich Orientierte Räume (LOR) als RDF (LOR V 01.01.2021)

Datensatz

Die lebensweltlich orientierten Räume (LOR) sind geografische Einheiten auf vier hierarchischen Ebenen: Bezirke, über 58 Prognoseräume, 143 Beobachtungsräume und 542 Planungsräume. Die LORs bilden die Basis für die Prognose und Beobachtung demografischer Veränderungen in Berlin. Dieser Datensatz bildet die Daten der LORs als Linked Data bzw. RDF ab.

Typ:	void:Dataset
zuerst erstellt am:	2022-12-31
zuletzt geändert am:	2023-03-28
Veröffentlichende Stelle:	BerlinOnline Stadtportal
Homepage:	https://github.com/berlinonline/lod-berlin-lor
Einstiegsressource:	lor:berlin
Beispielressource:	lor:plr_02500729
Beispielressource:	lor:pgr_0250
Beispielressource:	lor:bzr_025007
Beispielressource:	lor:bez_02
Andere Version:	lor19:
Datensatz-URL:	https://berlinonline.github.io/lod-berlin-lor/bzr_023004

Friedrichshain-Kreuzberg

[Datensatz](#) / [Berlin](#) / Friedrichshain-Kreuzberg

Untereinheiten (Prognoseräume)

- [Kreuzberg Nord](#)
- [Kreuzberg Süd](#)
- [Kreuzberg Ost](#)
- [Friedrichshain West](#)
- [Friedrichshain Ost](#)



Andere Version: [lor19:bez_02](#)

[Toggle Local Graph](#)

[Impressum](#)

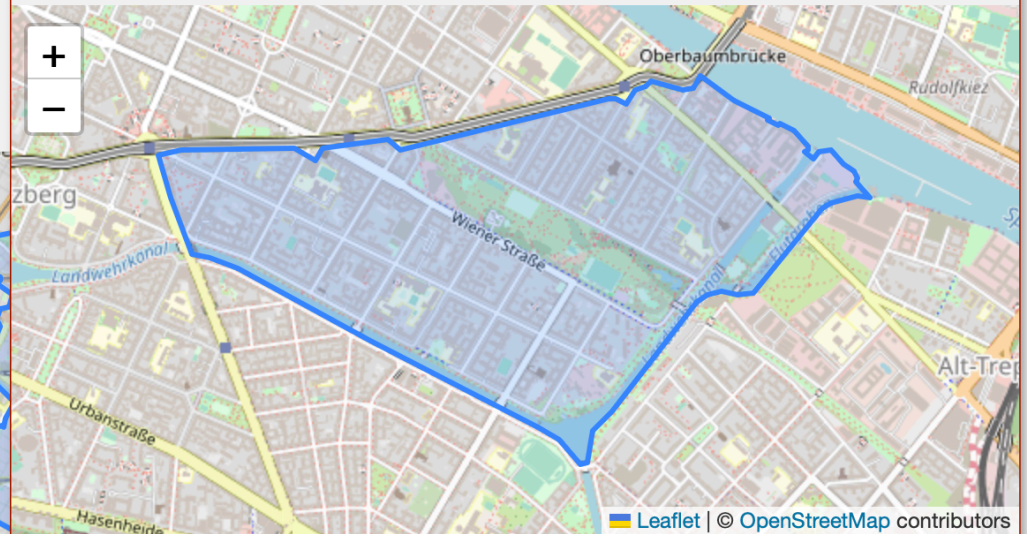
https://berlinonline.github.io/lod-berlin-lor/bez_02

Südliche Luisenstadt

[Datensatz](#) / [Berlin](#) / [Friedrichshain-Kreuzberg](#) / [Kreuzberg Ost](#) / Südliche Luisenstadt (Bezirksregion 023004)

Untereinheiten (Planungsraum)

- [Reichenberger Straße West](#)
- [Reichenberger Straße Ost](#)
- [Wrangelkiez](#)



Andere Version: [lor19:bzr_020304](#)

[Toggle Local Graph](#)

https://berlinonline.github.io/lod-berlin-lor/bzr_023004



BerlinOnline Stadtportal GmbH & Co. KG

[Datensatz](#) / BerlinOnline Stadtportal GmbH & Co. KG

Typ: [schema:Organization](#)

Webseite: <https://berlinonline.net>

Telefon: +49 30 2327 5530

Fax: +49 30 5771180 95

USt-IdNr: DE219483549

Adresse

Stefan-Heym-Platz 1

10367 Berlin

DE

[Toggle Local Graph](#)

[Impressum](#)

<https://berlinonline.github.io/lod-berlin-bo/berlinonline>



Toggle Local Graph

Datensatz

Statements

Graph

Typ:

Webseite

Telefon:

Fax:

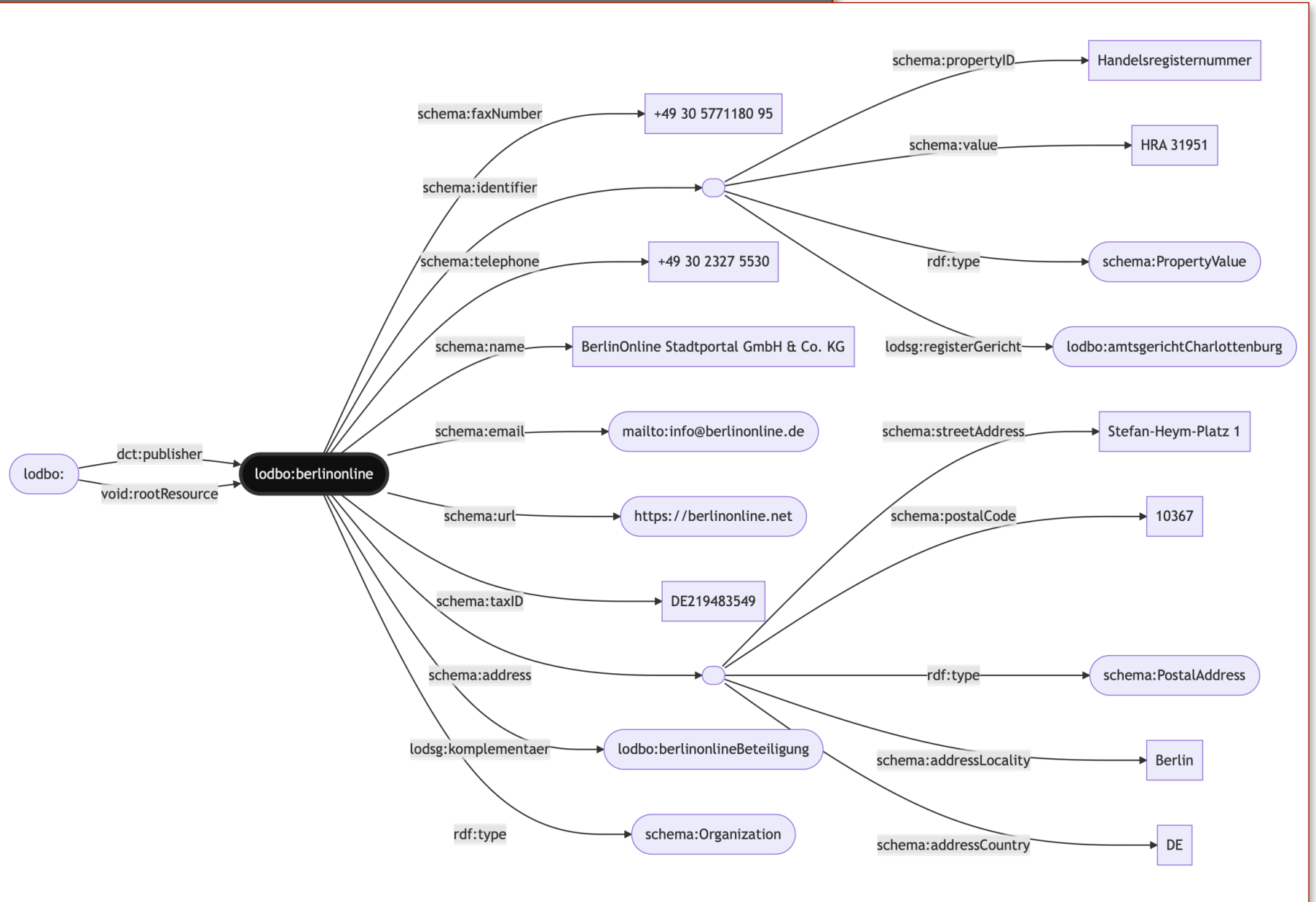
USt-IdN

Toggle

Impressum

Statements with [lodbo:berlinonline](#) as subject

Subject	Predicate	Object												
	schema:faxNumber	"+49 30 5771180 95"												
	schema:identifier	<table border="1"> <thead> <tr> <th>Predicate</th> <th>Object</th> </tr> </thead> <tbody> <tr> <td>schema:propertyID</td> <td>"Handelsregisternummer"</td> </tr> <tr> <td>schema:value</td> <td>"HRA 31951"</td> </tr> <tr> <td>rdf:type</td> <td>schema:PropertyValue</td> </tr> <tr> <td>lodsg:registerGericht</td> <td>lodbo:amtsgerichtCharlottenburg</td> </tr> </tbody> </table>	Predicate	Object	schema:propertyID	"Handelsregisternummer"	schema:value	"HRA 31951"	rdf:type	schema:PropertyValue	lodsg:registerGericht	lodbo:amtsgerichtCharlottenburg		
Predicate		Object												
schema:propertyID		"Handelsregisternummer"												
schema:value		"HRA 31951"												
rdf:type		schema:PropertyValue												
lodsg:registerGericht	lodbo:amtsgerichtCharlottenburg													
	schema:telephone	"+49 30 2327 5530"												
	schema:name	"BerlinOnline Stadtportal GmbH & Co. KG"												
	schema:email	mailto:info@berlinonline.de												
lodbo:berlinonline	schema:url	https://berlinonline.net												
	schema:taxID	"DE219483549"												
	schema:address	<table border="1"> <thead> <tr> <th>Predicate</th> <th>Object</th> </tr> </thead> <tbody> <tr> <td>schema:streetAddress</td> <td>"Stefan-Heym-Platz 1"</td> </tr> <tr> <td>schema:postalCode</td> <td>"10367"</td> </tr> <tr> <td>rdf:type</td> <td>schema:PostalAddress</td> </tr> <tr> <td>schema:addressLocality</td> <td>"Berlin"</td> </tr> <tr> <td>schema:addressCountry</td> <td>"DE"</td> </tr> </tbody> </table>	Predicate	Object	schema:streetAddress	"Stefan-Heym-Platz 1"	schema:postalCode	"10367"	rdf:type	schema:PostalAddress	schema:addressLocality	"Berlin"	schema:addressCountry	"DE"
		Predicate	Object											
		schema:streetAddress	"Stefan-Heym-Platz 1"											
		schema:postalCode	"10367"											
		rdf:type	schema:PostalAddress											
	schema:addressLocality	"Berlin"												
	schema:addressCountry	"DE"												
	lodsg:komplementaer	lodbo:berlinonlineBeteiligung												
rdf:type	schema:Organization													



[rdf:type](#)

[schema:Organization](#)

Repository Templates:

- **Vorlage** zum Erzeugen neuer Dataset-Repositories
- Fertige **Struktur**
- Fertige **Actions** zum Erzeugen der statischen Linked-Data-Site
- **HTML-Templates** für verschiedene RDF-Klassen
- Generische Templates (z.B. Anzeige aller lokalen Triples)
- LOD Site Generator: <https://github.com/berlinonline/lod-sg>

Einschränkungen



- **Skalierbarkeit:**
 - 100mb Dateigröße (aber: Git LFS ist möglich)
 - 25mb bei Browser-Upload
 - <5gb Repository-Größe ist “strongly recommended”
 - <https://docs.github.com/en/repositories/working-with-files/managing-large-files/about-large-files-on-github#about-size-limits-on-github>
- Kein **Content-Negotiation**, nur embedded JSON-LD
- Kein **SPARQL-Endpunkt**, nur statische Files
 - Daten runterladen, dann lokal (oder anderswo remote) abfragen

Aktuelle Beispiele

- Berliner LORs (ab 2021)
 - <https://berlinonline.github.io/lod-berlin-lor/>
- Berliner LORs (vor 2021)
 - <https://berlinonline.github.io/lod-berlin-lor-2019/>
- Demographische Daten zu Berlin
 - <https://berlinonline.github.io/lod-berlin-einwohner/>
- BerlinOnline
 - <https://berlinonline.github.io/lod-berlin-bo/>

Bildquellen

- **Clip-Art Quelle:** veröffentlicht von OpenClipart als Public Domain. <https://freesvg.org/water-spring>
- **Clip-Art Wüste:** veröffentlicht von JoyPixels unter Creative Commons Attribution 4.0 (CC BY 4.0): <https://creazilla.com/de/nodes/45457-wueste-clipart>
- **Clip-Art Handschlag:** veröffentlicht von JoyPixels unter Creative Commons Attribution 4.0 (CC BY 4.0): <https://creazilla.com/de/nodes/45390-handschlag-clipart>

Bildquellen

- **Clip-Art Sonnenstrahlen:** veröffentlicht von OpenClipart als Public Domain. <https://freesvg.org/golden-solar-rays>