

Using Linked Data for Monitoring Administrative Persecution Targeting LGBTQIA+ Expression

Andrei Nesterov^{1,*}, Sergey Katsuba¹

¹Dataout Foundation, The Hague, The Netherlands

Abstract

Evidence of human rights violations, such as discrimination based on sexual orientation and gender, is contained in various heterogeneous sources. Investigators aggregate it in text-based reports. However, it remains challenging to track the provenance of violation incidents and investigate connections between them. We address this in our use case, applying a Linked Data approach to monitor administrative judicial persecution targeting LGBTQIA+ expression in Russia. First, we collect data on such incidents from open sources, such as court documents and mass media publications. Second, we design an ontology that models the provenance information of the incidents and their content. Third, we build a knowledge graph that enables us to discover persecution patterns and systematically track violations. The ontology we developed can be applied to different types of monitoring independently of a legal context, and the knowledge graph can be extended to include incidents originating from different monitoring activities, organisations, and sources.

Keywords

Linked Open Data, Knowledge Management, Ontology Engineering, Knowledge Graph, Human Rights

1. Introduction and Background

Non-government organisations and the mass media play a crucial role in documenting and communicating human rights violations. Because their findings are often presented in reports, articles, and other textual formats, researchers have developed information extraction approaches to systematically analyse reported incidents [1]. However, even when data is extracted, these challenges remain: co-referencing incidents reported by different organisations, linking them to related entities and primary sources, and accessing provenance metadata. To address these challenges, we present a use case that applies a Linked Data approach to the monitoring of human rights violations.

Using Linked Data, we model incidents in which Russia's administrative law is applied to target LGBTQIA+ expression. After extracting data on such incidents, we develop an ontology and controlled vocabularies to build a knowledge graph (KG). Our modelling allows corroborating each incident with provenance information, linking sources that reported the same incidents, and linking entities from different incidents, such as defendants or content that triggered administrative proceedings. Such linking enables systematic tracking of persecution patterns: for example, when multiple administrative proceedings are initiated against the same individuals or companies.

In Russia, administrative legislation targeting LGBTQIA+ expression has been in force nationwide since 2013. Initially, such legislation prohibited disseminating information and performing public actions (such as protests) related to the so-called “*propaganda of non-traditional sexual relations*” among minors. In 2022, this legislation was extended to prohibit the “*propaganda*” overall, regardless of age. Another repressive mechanism emerged in 2023 after the Russian Supreme Court designated “*International LGBT Movement*” as an “extremist organisation”. “Participating” in such an organisation or displaying its symbols (for example, a pride flag) is an administrative and criminal liability [2].

We monitor instances, in which Russian courts punish individuals and organisations based on the administrative articles for “*propaganda of non-traditional sexual relations and/or preferences, sex change,*

Posters, Demos, Blue Sky, and Tutorials at SEMANTiCS 2026, Sep 2026, Ghent, Belgium

*Corresponding author.

✉ andrei.nesterov@dataout.org (A. Nesterov); sergey.katsuba@ucdconnect.ie (S. Katsuba)

ORCID 0000-0001-9697-101X (A. Nesterov); 0000-0001-8979-4211 (S. Katsuba)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

or refusal to have children” (Article 6.21) or for distributing such information among minors (Article 6.21.2).¹ Either article targets LGBTQIA+ expression. Examples are bookshops being fined for selling queer-themed literature or people being punished for wearing gender non-conforming clothing. These measures are discriminatory and violate human rights, such as freedom of expression.² Hence, we use the term “administrative persecution” (and not “prosecution” or “violation”): the administrative laws, discriminatory in their nature, are used by the state as an instrument of repression [3].

The issue of targeting LGBTQIA+ expression is not unique to Russia; it has also been observed in the European [4] and US contexts [5]. We aim to make our Linked Data approach applicable to representing persecution incidents originating from different legislative systems and sources.

In April 2026, the Judicial Department of the Supreme Court of Russia removed its statistics over 20 years and did not publish the data for 2025, as required by law.³ Researchers and journalists relied on this official data to report on state repression in Russia. In this context, it is especially crucial to make provenance-backed data on human rights violations accessible and reusable.

This work is a part of the Dataout Foundation’s project Grey Rainbow, which documents anti-LGBTQIA+ discrimination and hate crimes in Russia.⁴ We use Linked Data to systematise our monitoring and foster the reuse of our output.

2. Related Work

The Semantic Web technologies applied to the human rights domain can offer NGOs and investigators more effective methods of obtaining insights. Constructing a knowledge graph (KG) brought together heterogeneous sources of potential human trafficking incidents to enable investigations by NGOs and law enforcement [6]. In another work, researchers applied an approach similar to ours by extracting data from court cases on gender-based violence, developing an ontology, and building a KG to improve access to legal information [7]. Similarly, we apply a bottom-up approach. However, we represent not legal documents (for example, their structure or judges’ reasoning) but the provenance of incidents and their content with the goal of preserving and analysing them.

Previous research analysed Russian administrative court cases about LGBTQIA+ expression from a social science perspective [8]. Human rights organisations have also presented their reports on such cases [9]. However, these outputs did not include datasets with traceable incidents. Our goal is to construct a reusable KG which aggregates all known incidents with provenance information and enables their systematic tracking and analysis.

3. A Knowledge Graph of Persecution Incidents

After collecting data on persecution incidents, we develop a custom ontology and controlled vocabularies. Then, we construct a KG and validate it with competency questions. The resulting KG enables us to investigate connections between incidents and patterns of persecution. We outline ethical considerations that affected our data collection and KG development processes.

3.1. Data

We have documented 316 incidents of administrative persecution targeting LGBTQIA+ expression by extracting data from open sources. The primary source was court documents published on the official court websites. We retrieved such documents dated from 2014 to 2025. Among other sources

¹According to the Code of Administrative Offences of the Russian Federation

²See, for example, the judgement of the European Court of Human Rights in the Case of Bayev and others v. Russia: <https://hudoc.echr.coe.int/eng?i=001-174422>

³The Moscow Times. “Russia Removes 2 Decades of Court Statistics From Public Access”, 21.04.2026: <https://www.themoscowtimes.com/2026/04/21/russia-removes-2-decades-of-court-statistics-from-public-access-a92561>. Note: As of the end of June, access to the statistics has been restored; however, the new data for the full year of 2025 has not been released.

⁴<https://greyrainbow.dataout.org>

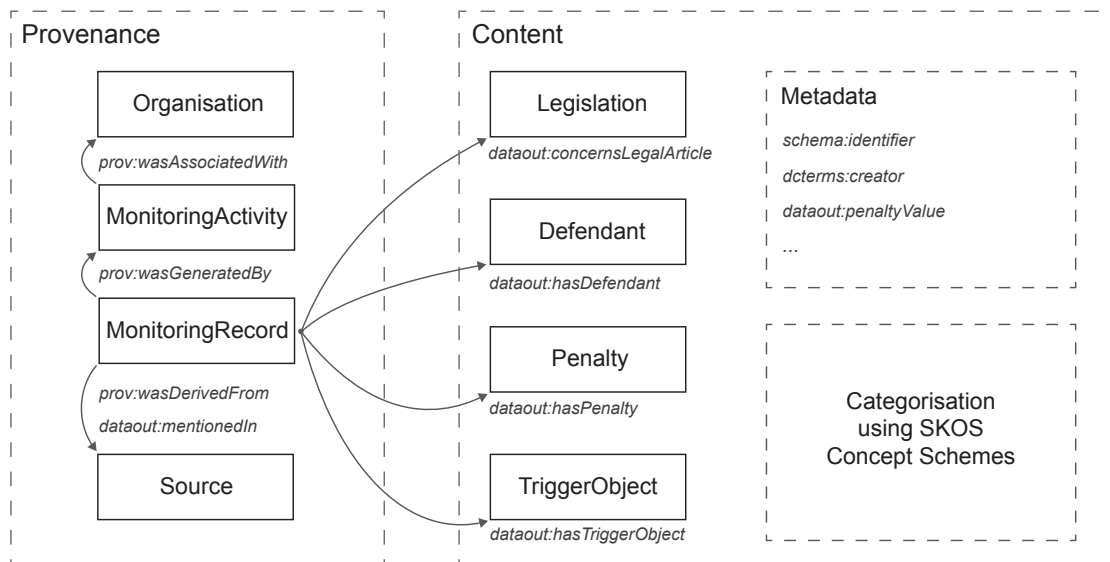


Figure 1: Representing provenance and content of monitoring records with custom and reused solutions.

were courts' press releases about initiated proceedings and publications from mass media and social media. The latter originate from representatives of courts or law enforcement. Sources other than court documents provided additional information in cases when courts redacted important details, such as fine amounts. In the current work, we collected only cases which resulted in penalties and for which we obtained sufficient evidence.

Among the core information we extracted from the documents are the administrative article applied, information or actions that triggered persecution (we use the term *trigger objects*), a defendant, penalty imposed, and a case summary. The information extraction was carried out manually: both authors read each document to ensure accuracy and prevent the publication of personal data contained in documents. Some documents' metadata was collected automatically, such as publication date and ID. In the dataset, we provide extracted textual information in English and Russian. Where possible, we linked incidents to external sources: media publications and trigger objects. If an incident was mentioned in Russian- or English-language mass media, we provided archived URLs to such mentions obtained through the Internet Archive's Wayback Machine.⁵ We also linked trigger objects to resources about them: for example, all films and TV series which triggered courts to issue penalties are linked to their IDs from The Movie Database (TMDB).⁶ For each incident, we preserved its provenance information.

3.2. Developing Ontology and Controlled Vocabularies

We developed the Dataout Ontology to represent incidents of persecution resulting from our monitoring.⁷ Figure 1 presents how we modelled the provenance information and content of incidents.

Our approach is provenance-centric: we represented incidents as phenomena we observed in sources. We applied a W3C recommendation, the PROV Ontology (PROV-O),⁸ to model core provenance classes and properties. Each incident we observe is an instance of *MonitoringRecord* (subclass of *prov:Entity*), which is a product of *MonitoringActivity* (*prov:Activity*) carried out by *prov:Organization*. This design allows representation of incidents identified during different monitoring activities by different organisations. Each monitoring record is corroborated by at least one *Source* (*prov:Entity*). A single incident may be mentioned in multiple sources, for example, when the media reports about a court case. For such a connection, we introduced the property *mentionedIn*.

⁵<https://web.archive.org>

⁶<https://www.themoviedb.org>

⁷Dataout Ontology: <https://dataout.org/ontology>

⁸<https://www.w3.org/TR/prov-o/>

To model incidents' content, we introduced custom classes and properties if reusing the existing ones did not correspond to our data. The content of an incident primarily consists of facts found in a court document: the administrative article applied (we reused *schema:Legislation*), a defendant (against whom persecution is carried out), penalty imposed by the court, and what triggered the authorities to initiate the proceedings (*TriggerObject*). The latter can be an information object (such as a social media post) or a public action (such as a single protest), for which companies or individuals were persecuted. In total, there are seven custom classes. We introduced 18 custom properties, which can be divided into three groups. The first connects instances of different classes: a monitoring record *hasTriggerObject*. The second connects instances to their categories: a trigger object has *triggerObjectCategory*. And the third group is data properties, such as *uniqueCourtDocumentID*, a sub-property of *schema:identifier*, which connects a court document, an instance of *Source*, to its ID assigned by the court. The incidents' metadata is represented using Schema.org⁹ and DCMI Metadata Terms.¹⁰

We compiled six controlled vocabularies using the SKOS Concept Scheme.¹¹ They are intended to categorise instances related to the provenance (Monitoring Activity, Source) and content of incidents (Legislation, Defendant, Penalty, and Trigger Objects). Using the controlled vocabularies, we avoid inflating our ontology with many (non-domain-specific) classes. For example, we identified 14 categories of trigger objects, so we constructed *TriggerObjectCategoryScheme* with 14 concepts. These concepts represent media items (books, films, music videos, etc.), web content (social media posts, dating profiles, etc.) and public actions that the authorities deemed as violations.

Where possible, we identified matches between the concepts in our controlled vocabularies and external concepts using the *related*, *close*, and *exact* match SKOS properties. Some concepts were connected with the *broader* and *narrower* relationships. For example, the concept "Protest" from the *TriggerObjectCategoryScheme* has a close Wikidata match¹² and the broader concept "Public action". The custom controlled vocabularies are documented on the same page as the ontology.

3.3. Constructing and Querying Knowledge Graph

We assigned URIs to the instances in the monitoring data that we initially collected in a tabular format: incidents, administrative articles, defendants, penalties, sources, trigger objects, courts, and regions. Using the developed ontology and controlled vocabularies, we converted the tabular data into a KG with RDFLib¹³ in Python. The resulting KG has been documented and made openly available on our website in several serialisations: JSON-LD, RDF/XML, N Triples, and TTL.¹⁴

We evaluated the KG with competency questions (CQ).¹⁵ The ten CQs we formulated in SPARQL check for missing data, count and aggregate instances. For example, in one CQ, we grouped penalties by type (fine, detention, expulsion, warning) and checked their distribution by year. We confirmed that the query results corresponded to the results of filtering the tabular data.

Using the KG enables us not only to preserve the monitored incidents with their provenance and related sources in a structured manner, but also to investigate connections between incidents and identify patterns of persecution. For the latter, we constructed analytical SPARQL queries.¹⁵ We analysed the distribution of incidents by year, region, administrative article, and category of punishment. Having URIs for each trigger object, we determined which appeared in multiple incidents: generally, these are cases about different streaming platforms fined for the same film. We also established for which categories of trigger objects different defendants were persecuted. Grouping incidents by defendants and trigger objects, we have found that individuals were persecuted mainly for posting on social media, while legal entities and their representatives were persecuted for films, TV series, comics, and books. In cases with defendants-individuals, we observed "serial" persecution, when multiple proceedings are

⁹<https://schema.org/>

¹⁰<https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>

¹¹<https://www.w3.org/TR/skos-reference/>

¹²<https://www.wikidata.org/wiki/Q273120>

¹³<https://github.com/RDFLib/rdflib>

¹⁴<https://dataout.org/dataset/project-6-21>

¹⁵See the project's GitHub repository: <https://github.com/dataout-org/project-6-21>

initiated against the same individual. Selecting and grouping incidents by defendants, we have found 16 instances (59 court cases) in which individuals were persecuted more than once. For example, one social media user was fined nine times for posting images online.

Investigating other connections requires non-trivial queries. For example, one of the patterns we observed while collecting data is when legal entities and their affiliated individuals were persecuted because of the same trigger object. With affiliation links extracted from court documents, we can reveal this pattern with a SPARQL query. This query includes two sub-queries: the first selects trigger objects from incidents in which a defendant is affiliated with a company; the second selects incidents with these companies acting as defendants and the same trigger objects. The result of this query is 39 instances (78 court cases) of when companies and their employees were fined for the same “violation”. For example, both the legal entity of a Moscow bookshop and its director were fined for selling five titles related to LGBTQIA+ and feminism. By querying the KG, we can produce insights into how the anti-LGBTQIA+ administrative law is being applied and collect evidence of the persecution patterns.

3.4. Ethical Considerations

Our work deals with sensitive and political issues. We considered potential risks and ethical moments during the data collection, ontology and knowledge graph development processes.

When collecting data from court documents, we encountered personal information. We do not include such information in our dataset. We include individuals’ names only when they serve as company representatives or are mentioned in the media, for example, activists or celebrities.

The language used in court documents and other sources, from which we collected the incidents, is stigmatising. In our dataset, we provide snippets of incidents which contain quotes with stigmatising phrases; we do this to illustrate the position of judges and judicial experts.

Because of the discriminatory nature of the laws the Russian courts used to target LGBTQIA+ expression, we refrained from modelling the legality of court cases when developing the ontology. For this reason, we did not reuse existing legal vocabularies or ontologies. For example, instead of representing “violations”, we introduced the class *TriggerObject* to indicate what judges interpreted as a violation. In the KG, we link all incidents to the activity of monitoring administrative persecution.

4. Conclusion

Using Linked Data, we represented the results of our monitoring of administrative persecution targeting LGBTQIA+ expression in Russia. We built a knowledge graph of such persecution incidents by applying a bottom-up approach: first, we gathered data from open sources, and second, we developed an ontology and controlled vocabularies. The ontology can be reused for representing the outcomes of other monitoring. The knowledge graph enabled us to represent evidence of human rights violations in a machine-readable format, supporting it with provenance information and links to external sources. It also facilitates the discovery of connections between incidents and patterns of persecution. In future work, we will extend the knowledge graph with new incidents, as we uncover them.

Author Contributions

We list author contributions according to the Contributor Roles Taxonomy (CRediT). Conceptualization: AN, SK; Data curation: AN, SK; Methodology: AN, SK; Visualization: AN; Writing (original draft): AN.

Declaration on Generative AI

During the preparation of this work, the authors used Grammarly in order to: check spelling and punctuation. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication’s content.

References

- [1] C. Djouvas, N. Ioannidis, I. Kovras, C. Christodoulou, Human rights violations reporting dataset, Data in Brief 66 (2026) 112854. URL: <https://www.sciencedirect.com/science/article/pii/S2352340926004063>. doi:10.1016/j.dib.2026.112854.
- [2] Human Rights Watch, Russia: Rising Toll of LGBT 'Extremism' Designation, Online, 2025. URL: <https://www.hrw.org/news/2025/06/30/russia-rising-toll-of-lgbt-extremism-designation>.
- [3] E. Bapt, R. Sandford, Invisible Citizens: How Targeting And Repression of Sexual Orientation and Gender Identity Minorities in Russia Impacts Human Rights, Online, Geneva Academy of International Humanitarian Law and Human Rights, 2025. URL: <https://geneva-academy.ch/wp-content/uploads/2026/01/How-Targeting-and-Repression-of-SOGI-Minorities-in-Russia-Impacts-Human-Rights.pdf>.
- [4] N. Gupta, State-Sponsored Homophobia: In Relation with the New Hungary Laws, 2021, International Journal of Law Management and Humanities 4 (2021) 405 – 417. URL: <https://ijlmh.com/paper/state-sponsored-homophobia-in-relation-with-the-new-hungary-laws-2021/>.
- [5] C. A. Ball, "Harmful to Children" Claims and the Targeting of LGBTQ Speech *After Free Speech Coalition v. Paxton*, Yale Law Journal Forum 135 (2026). doi:10.2139/ssrn.5736803.
- [6] P. Szekely, C. A. Knoblock, J. Slepicka, A. Philpot, A. Singh, C. Yin, D. Kapoor, P. Natarajan, D. Marcu, K. Knight, D. Stallard, S. S. Karunamoorthy, R. Bojanapalli, S. Minton, B. Amanatullah, T. Hughes, M. Tamayo, D. Flynt, R. Artiss, S.-F. Chang, T. Chen, G. Hiebel, L. Ferreira, Building and Using a Knowledge Graph to Combat Human Trafficking, in: M. Arenas, O. Corcho, E. Simperl, M. Strohmaier, M. d'Aquin, K. Srinivas, P. Groth, M. Dumontier, J. Heflin, K. Thirunarayan, S. Staab (Eds.), The Semantic Web - ISWC 2015, Springer International Publishing, Cham, 2015, pp. 205–221. doi:10.1007/978-3-319-25010-6_12.
- [7] C. D'Amato, G. Rubini, F. Didio, D. Francioso, F. Z. Amara, N. Fanizzi, Automated Creation of a Legal Knowledge Graph Addressing Cases of Violence Against Women. Resource Methodology and Lessons Learned, in: R. Markovich, L. Di Caro, A. Rapp, C. Schifanella (Eds.), Frontiers in Artificial Intelligence and Applications, volume 416, IOS Press, 2025, pp. 37 – 48. doi:10.3233/FAIA251575.
- [8] S. Katsuba, Russian "Gay Propaganda Law": A Comprehensive Qualitative Analysis of the Legislation and Case Law, Problems of Post-Communism 73 (2026) 138–151. doi:10.1080/10758216.2025.2487577.
- [9] Citizens' Watch and Sphere Foundation, Judicial harassment against LGBT+ under 'propaganda' law in Russia, Online, 2024. URL: <https://spherequeer.org/judicial-harassment/>.