

Open Data Foraging: a Semantic Approach to Findability and Understandability

Eudes Adiba^{1,*}, Antoine Clarinval² and Anthony Simonofski¹

¹Namur Digital Institute, University of Namur, Rue de Bruxelles 61, 5000 Namur, Belgium

²Swedish Center for Digital Innovation, Department of Informatics, Umeå University, Umeå, Sweden

Abstract

Findability and understandability are two main barriers to Open Government Data (OGD) reuse. Instead of addressing these barriers at the individual dataset level, we present a shift of perspective in their definition through Information Foraging Theory and propose a semantic annotation methodology to solve these problems by reshaping the information space and improving the quality of information scent from each dataset.

Keywords

Open Government Data, Semantic annotation, Information Foraging Theory

1. Introduction

The potential benefits associated with effective OGD reuse face two persistent obstacles: the difficulty of finding relevant datasets (findability) and the difficulty of understanding them sufficiently for reuse (understandability) [1]. Current approaches often address these issues by improving metadata at the individual dataset level. While these efforts are necessary, they are not sufficient. Exploring open data is not simply a one-off search operation [2]. It is a dynamic search process, made of trial and error, forking and backtracking, in which the user progressively develops an understanding of the information territory (OGD portals). Given this context, the important thing is not just the isolated quality of the (meta)data, but the way in which the data environment guides or disorients navigation. This shift of perspective leads to the need for a redefinition of the two problems not as properties of the datasets themselves but as experiences, shaped by the structure of the information space and the signals it emits. To model this perspective, the current work draws on Information Foraging Theory (IFT) to redescribe dataset search on OGD portals and proposes a semantic annotation designed as an information overlay to solve the two obstacles by: (1) reshaping the information space and (2) improving the quality of information scent from each dataset. With this poster, we aim to give an overview of this new perspective informed by IFT and the semantic annotation methodology. The goal is to engage in discussions on how these can (re)shape the OGD community's work.

2. IFT Perspective and Contribution

2.1. OGD Search from IFT Perspective

Information Foraging Theory (IFT) [3] is about how people seek information derived from animal food-foraging strategies. It has successfully explained people's information seeking behaviors in various domains and is based on a set of constructs (Table 1). IFT models information (dataset) seeking as an adaptive behavior, where the citizen (OGD portal user) explores an information environment (the OGD portal) in search of prey (relevant dataset). This environment is made up of patches (thematic collections) connected by links (hyperlinks, tags, suggestions), and each link provides cues. The information scent corresponds to the user's estimate, based on these cues, of the potential usefulness of the dataset at the other end of the link. Through this mechanism, information scent influences the browsing choices between the patches.

Proceedings EGOV-CeDEM-ePart conference, August 31 - September 4, 2025, University for Continuing Education, Krems, Austria.

*Corresponding author.

✉ eudes.adiba@unamur.be (E. Adiba)

ORCID 0009-0000-6316-9100 (E. Adiba); 0000-0003-3936-0532 (A. Clarinval); /0000-0002-1816-5685 (A. Simonofski)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Table 1
IFT Constructs

Construct	Definition	Example in OGD search
Predator	Person foraging for information	Citizen / OGD Portal user
Prey	Information that the predator is seeking	Relevant dataset
Information environment	Environment where the foraging happens	OGD Portal
Patch	A cluster of prey	A thematic group of datasets
Link	A navigational element that connects patches	Hyperlinks based on similar topics, tags, or linked metadata
Cues	Small details that help predator guess what they'll find if they follow a link	Metadata
Scent	Predator's perception of the value of a link based on cues	Perceived usefulness of a dataset

2.2. OGD Semantic Annotation

OGD portals often organize data into collections that correspond to patches. These patches are based on tags which, due to their inconsistency, can lead to patches with heterogeneous content: an aspect that can confuse users in their searching process. Moreover, users must rely on the portal's predefined structure to navigate, as there are no clear links within or between patches. This is tedious and adds to the cognitive load of the user. Another key point is the information scent emanating from the metadata, which helps the user judge the relevance of a dataset. The metadata use different terminologies depending on the data provider; this greatly weakens the scent.

Semantic annotation (the process of identifying and linking the real-world concepts of a knowledge graph to different elements of tabular data) makes it possible to structure patches more coherently by leveraging the semantic relations of associated concepts. Using the relations of the knowledge graph, one can add meaningful links within and between patches. This also helps to increase information scent, as the cues provided are based on concepts and terminologies familiar to users. We proposed a probabilistic approach (Conditional Random Field) to OGD annotation task by integrating the diversity of data type and the richness of OGD structure information: header, metadata, values in the process (see Figure 1).

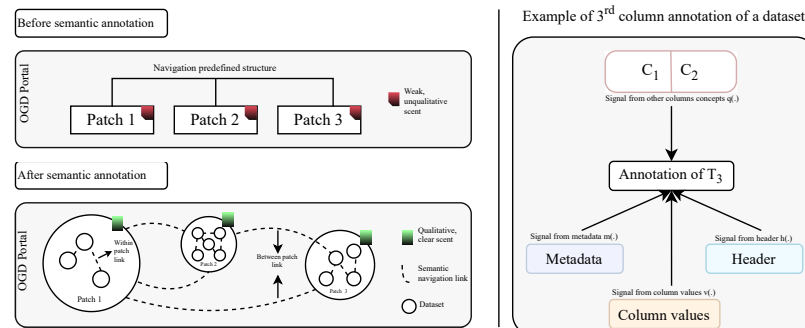


Figure 1: Process of data annotation

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

- [1] P. Kremen, M. Necasky, Improving Discoverability of Open Government Data with Rich Metadata Descriptions Using Semantic Government Vocabulary, 2018.
- [2] J. R. Crusoe, K. Ahlin, Users' activities for using open government data – a process framework, Transforming Government: People, Process and Policy 13 (2019) 213–236.
- [3] P. Pirolli, S. Card, Information foraging, Psychological Review 106 (1999) 643–675.