

# The impact of Grey Literature in the web environment: A citation analysis using Google Scholar

Rosa Di Cesare, Daniela Luzi, Roberta Ruggieri

Consiglio Nazionale delle Ricerche, Istituto di Ricerche sulla Popolazione e le Politiche Sociali

## 1. Introduction

The use of Grey Literature (GL) has hitherto been studied on the basis of whether and to what extent GL documents had been cited in peer-reviewed conventional literature applying citation analysis techniques based on primary sources (i.e. bibliographic references of journal articles) [Alb2000, Dic2004], or on the multidisciplinary citation indexes produced by Thomson ISI, the Web of Science (WoS) [Pel2003, Cor2004].

More recently, other tracking citation systems have been developed, such as Scopus, Google Scholar (GS), Citeseer, and CrossRef. To some extent these have challenged the ISI monopoly over scientific evaluation of countries, institutions, groups and authors. Accordingly, many studies have begun to compare the different systems, particularly WoS, Scopus and GS [Bak2006, Jac 2005, Nor2005], to reveal the citations they have in common and the discrepancies, both in terms of frequency and source. In the process, the limits of the individual systems have emerged. These include citation errors (typographical errors, non-standard reference formats, parsing errors, etc.), as well as varying focus according to discipline and the varying document types used for tracking citation counts.

Our research concentrates principally on the latter aspect and aims to ascertain the impact of GL on the web environment. Consequently we chose to use GS that considers not only citations from peer-reviewed conventional journals, but also includes citations received by GL documents. Potentially, GS is the appropriate tool for illustrating the various changes underway in scholarly communication as it can reveal not only different electronic documents, predominantly GL documents, to be found in the assortment of institutional pages, open archives and repositories, but can also reveal their impact in terms of the citations made. Therefore, a secondary aim of this paper is to verify whether GS is an efficient tool to identify core papers as well as in tracking citations from different document types and whether it is able to represent the scholarly communication deriving from citations considering both GL and conventional literature. This is fundamental for GL, which is clearly an integral part of scholarly communication for numerous scientific disciplines, but still remains on the periphery of research evaluation activities.

The paper presents the results of a citation analysis with the aim of evaluating the use of GL through a comparison of the number of citations received by both GL and conventional documents. The analysis is carried out retrieving documents dealing with the topic *population ageing*, a "hot multidisciplinary topic", which is studied under demographic, socio-economical or medical perspectives.

## 2. Methodology

We have chosen to carry out the citation analysis using GS because this tracking citation index is:

- The only open source system, freely available on the Internet,
- It claims to have a multidisciplinary subject coverage especially in social sciences and humanities, and last but not least
- It tracks citations to GL.

Ageing is particularly suited to citation analysis as it has been a hot topic in recent years. Research on ageing has been carried out in our Institute, where authoritative bibliographic material is available. Moreover, many governmental and intergovernmental institutions carry out research on this topic under various perspectives, (demographic, socio-economic and health aspects) and they generally produce and diffuse their results using GL documents. [Dic 2006].

Our study is divided into two parts. The first one analyses the cited documents retrieved in GS in the topic chosen. We define as *cited documents* the documents that have received at least 10 citations. From the cited documents we selected a set of items, which have received at least 50 citations and we define them *highly cited documents*. In the second part we identified the citing documents of the highly cited ones focusing in particular on the relationship between cited and citing documents in terms of publication type and citation over-time.