Article

Evaluating Open Access Journals using Web Semantic Technologies and Scorecards

Hallo Maria
National Polytechnic School (ECUADOR)

Luján-Mora Sergio
University of Alicante (SPAIN)

Maté Alejandro
University of Alicante (SPAIN)

Abstract
This paper describes a process to develop and publish a scorecard from an OAJ (Open Access Journal) on the semantic web using Linked Data technologies in such a way that it can be linked to related datasets. Furthermore, methodological guidelines are presented with activities related to each step of the process. The proposed process was applied to a university OAJ from a university, including the definition of the KPIs (Key Performance Indicators) linked to the institutional strategies, the extraction, cleaning and loading of data from the data sources into a data mart, the transformation of data into RDF (Resource Description Framework), and the publication of data by means of a SPARQL endpoint using the Virtuoso software. Additionally, the RDF data cube vocabulary has been used to publish the multi-dimensional data on the Web. The visualization was made using CubeViz, a faceted browser to present the KPIs in interactive charts.

Keywords
Linked Data; semantic web, RDF data cube vocabulary, knowledge management.

1. Introduction

OA (Open Access) is the free unrestricted online access to digital content. OAJ (Open Access Journals) are scholarly journals that are available online to the reader "without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself" [1]. A classification suggested by Suber [2] for the OA content based on the rights that authors keep to disseminate their work is summarized as follows:

- Gold Open Access is adopted by peer reviewed journals, making the published version freely available from the publisher's server without any other rights or permissions being granted.
- Green Open Access allows authors for self-archiving in repositories with the consent of journal or publishers. These repositories are discipline specific or institutional.
- Pale Green allows authors to archive preprints.
- Gray allows authors make their work accessible on institutional or personal websites.

Corresponding author:
María Hallo, National Polytechnic School, Isabel la Católica E11-253, Quito, Ecuador PO-Box 17-01-2759
Email: maria.hallo@epn.edu.ec

The last version of this article was published online before print, January 13,2016, Journal of Information Science, 2016, © The Author(s), DOI: 10.1177/0165551515324353,

http://jis.sagepub.com/content/early/2016/01/13/0165551515624353.abstract

In another proposal [3], OAJ are classified as traditional, pure open access and hybrid:

- Traditional subscription-based journals charge annual subscription fees and deliver their content to subscribers only.
- Pure open access journals make all articles available for free online immediately on publication.
- Hybrid journals are subscription journals which offer an option for immediate open access for individual articles. The authors have the option to pay to provide OA to everybody.

A growing number of scholarly journals are using OJS (Open Journal System), a software platform designed to manage articles through author submission, the peer review process, edition and publication [4, 5]. While such system fosters the publication process, little attention has been paid to evaluate the use of OAJ.

OAJ routinely collect statistics about the use of their digital collection for evaluation purposes. However, these statistics are dispersed, stored across repository files lacking a standard structure, and unrelated to the business objectives. As a result, it is difficult for researchers and users to compare statistical information, while for OAJ it becomes a challenge to develop policies, assess the impact of its use in society, and share their discoveries.

In order to tackle this problem, this paper proposes a scorecard, a tool to monitor strategic objectives in a business [6], for evaluating and comparing OAJ based on statistics suggested in the ISO 2789:2013 standard [7], as well as a technical architecture for publishing them based on Linked Data technologies. The term Linked Data refers to a set of best practices for publishing and interlinking structured data on the Web in a human and machine readable way [8].

The proposed approach for evaluating the use of OAJ using Linked Data technologies was developed based on best practices and recommendations from several authors [9, 10] and tested with a case study based on the journal "Revista Politécnica"[1], in the context of a interuniversity project for publishing library bibliographic data using Linked Data technologies, developed by National Polytechnic School from Quito (Ecuador) and other universities. "Revista Politécnica" is a scientific OAJ whose data were used to demonstrate the value of the proposed linked open data analytics approach. The dataset contained metadata of scientific articles published in 2014 under an open license. In addition, the dataset created was linked to external data providing information that goes far beyond the bibliographic data supplied by publishers, such as number of papers in similar subjects, number of visits, statistical indicators below national standards, etc. The results of these evaluation strategies can have a number of significant implications for the continued development and improvement of OAJ.

The remainder of this paper is structured as follows. Section 2 presents the background on Linked Data technologies describing the principles and more important characteristics of vocabularies and formats. Section 3 describes scorecards and the metrics used for evaluating OAJ. Section 4 presents our proposal for defining and publishing a scorecard for the evaluation of OAJ using RDF formats. Finally, Section 5 describes the conclusions and sketches future works.

2. Background

Following we present some concepts used in this work: Linked Data, URIs, RDF, OWL and multidimensional data models.

2.1. Linked Data

The term Linked Data refers to a set of best practices for publishing and interlinking structured data on the Web in a human and machine readable way [8]. It is based on the URI (Uniform Resource Identifier)[2] and RDF (Resource Description Framework) specifications[3] . The Linked Data principles are:

- Use URIs as names for things.
- Use HTTP URIs so that people can look up those names.
- When someone looks up a URI, provide useful information using of entities and/or relations from a data model using controlled vocabulary terms and common standards such as RDF and SPARQL (RDF query language).
- When someone looks up a URI, provide useful information, using common standards such as RDF (Resource Description Framework) and SPARQL (RDF query language).
- Include links to other URIs so that they can help to discover related data.

The last version of this article was published online before print, January 13,2016, Journal of Information Science, 2016, © The Author(s), DOI: 10.1177/0165551515324353,

http://jis.sagepub.com/content/early/2016/01/13/0165551515624353.abstract

2.2. Naming things with URIs

In Linked Data, the items in a domain of interest and their relations are identified by HTTP URIs. An HTTP URIs should be dereferenceable helping clients to retrieve a description of the resource that is identified by the URI. The document Cool URIs for the Semantic Web[2] presented by W3C Interest Group describes strategies to make URIs dereferenceables.

2.3. RDF data model

RDF is used for publishing Linked Data on the Web, modelling and representing information resources as structured data.

In RDF, the fundamental unit of information is the triple (subject, predicate, object), a type of sentence that represent a simple fact about a resource.

Figure 1 shows graphically the structure of a RDF triple and two examples of RDF triples with the creator and title of an article:

- http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf, dc: creator, "Tim Berners-Lee"
- http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf, dc: title, "Linked Data – The Story So Far"
- dc: is an abbreviation for http://purl.org/dc/elements/1.1/ which means that dc:creator and dc:title are labels defined at this http address.

In each triple, the "subject" denotes the resource being described and it is represented by a URI. The "predicate" denotes a property of the subject or a relation between the subject, and the object. The predicate is generally a term from a well-known vocabulary or ontology represented by a URI. The "object" denotes the value of a property or another resource which is the target of the relation. Objects can be literals or URIs.

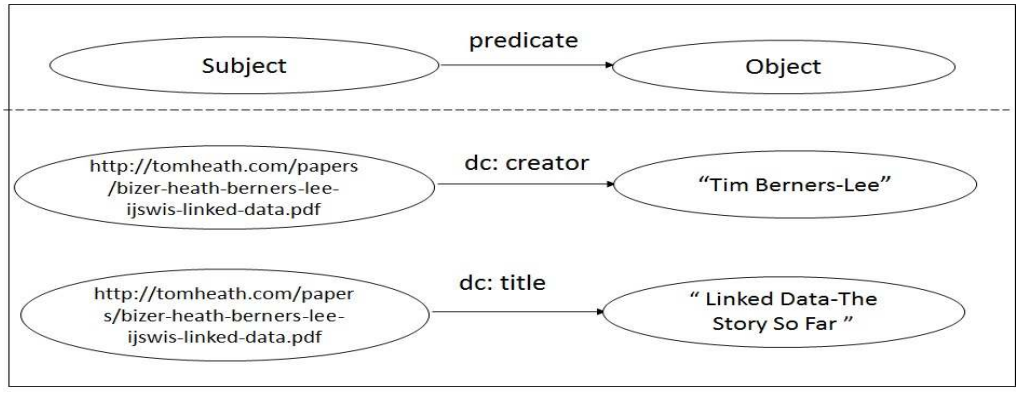

Figure 1. Examples of RDF triples.

Triples can be represented in different formats. For example, Figure 2 describes the triples from the Figure 1 in RDF/XML, a syntax defined by W3C to express an RDF graph as an XML document[4].

The last version of this article was published online before print, January 13,2016, Journal of Information Science, 2016, © The Author(s), DOI: 10.1177/0165551515324353,

http://jis.sagepub.com/content/early/2016/01/13/0165551515624353.abstract

```
<?xml versión= “1.0”?>
<rdf:RDF
xmlns:rdf= “http://www.w3.org/1999/02/22-rdf-syntaxis-ns# ”
xmlns:dc= “http://purl.org/dc/elements/1.1”>
  <rdf:Description rdf:about=“http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf”>
    <dc:creator> Tim Berners-Lee </dc:creator>
     <dc:title> Linked Data - The Story So Far </dc:title>
  </rdf:Description>
</rdf:RDF>
```

Figure 2. RDF/XML document.

Using a combination of URIs and RDF, it is possible to give identity and structure to data. However, using only these technologies, it is not possible to add semantics to data. Ontologies are used to provide semantics to data. An ontology represents knowledge as a hierarchy of concepts within a domain, using a shared vocabulary to denote the types, properties and interrelationships of those concepts.

2.4. RDFS and OWL

The Semantic Web Architecture includes two technologies: RDFS (RDF Schema) and OWL (Web Ontology Language). RDFS is an extension of RDF that defines a vocabulary for the description of entities and relationships[5]. RDFS describes subclass hierarchy and properties hierarchy. Some elements of the RDFS vocabulary are defined in Table 1, such as rdfs:Class, rdfs:resource, rdfs:subclassOf. In addition, RDFS adopts a property centric approach, the properties are defined in terms of the classes of resources to which they apply using rdfs:range and rdfs:domain which are instances of rdf:Property. This schema allows anyone to extend the description of existing resources. For example, we could define an eg vocabulary with eg:coauthor property, the domain eg:article and the range eg:person. Afterwards, anyone could define additional properties with the same domain and range using this RDF property centric approach.

Table 1. Core Class and Properties of the RDF Schema Vocabulary

Elements	Comment
rdfs: Resource	The class resource, everything
rdfs: Class	The class of classes
rdfs:Literal	The class of literal values,
rdf:Property	The class of RDF properties
rdfs:Datatype	The class of RDF datatypes
rdf:type	Instance of rdf:Property used to state that a resource is an instance of a class
rdf:subClassOf	Property used to state that the instances of one class are instances of another
rdfs:range	A range of the subject property
rdfs:domain	A domain of the subject property
rdfs:label	A human-readable version of a resource´s name
rdfs:comment	A human-readable description of a resource

In Figure 3, we present an example of RDF/XML document stating that an article is a class and a subclass of a document class.

In Figure 4, we have an example of a RDF/XML document stating that author is a property of article and takes literals as values.

The last version of this article was published online before print, January 13,2016, Journal of Information Science, 2016, © The Author(s), DOI: 10.1177/0165551515324353,

http://jis.sagepub.com/content/early/2016/01/13/0165551515624353.abstract

```
<?xml versión= "1.0"?>
<rdf:RDF
xmlns:rdf = "http://www.w3.org/1999/02/22.rdf.syntax.ns#"
xmlns:rdfs= "http://www.w3.org/2000/01/rdf-schema#">
   <rdfs:Class rdf:ID="article">
      <rdfs:subclassOf rdf:resource= "#document"/>
   </rdfs:Class rdf:ID="article">
</rdf:RDF>
```

Figure 3. An example of RDF/XML document using RDFS class elements.

```
<?xml versión= "1.0"?>
<rdf:RDF
xmlns:rdf = "http://www.w3.org/1999/02/22.rdf.syntax.ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">
   <rdf:Property rdf:ID="author">
      <rdfs:domain  rdf:resource= "#person"/>
      <rdfs:range rdf:resource="Literal"/>
    </rdf:Property>
</rdf:RDF>
```

Figure 4. An example of RDF/XML document using the rdfs:range and rdfs:domain elements.

OWL is an extension of RDFS[5] [8], in the sense that it uses class and properties providing additional metadata terms for the description of ontologies giving efficient reasoning support. Ontologies are formalized vocabularies of terms covering a specific domain and shared by a community of users. A set of concepts (e.g. entities, attributes, and processes), their definitions and their inter-relationships [11] are defined in ontologies. They are primarily exchanged as RDF documents and could be used along with information written in RDF. The first version of OWL Web Ontology Language was liberated in 2004 as a recommendation from W3C OWL working group. A new version the OWL[6] language was published in 2012. In the latest version, any OWL 2 ontology can also be viewed as an RDF graph.

2.5. Multidimensional Data Model and RDF

In addition to the standards described in the previous sections, it is necessary to describe the KPIs (Key Performance Indicators), measurable values that demonstrate how effectively a company is achieving key business objectives [12], in a multidimensional data model in order to enable its analysis, with this purpose we use the RDF Data Cube vocabulary to publish, discover, and link statistical data. The multidimensional data model has dimensions. A dimension represents a business perspective under which data analysis is to be performed and is organized in a hierarchy of levels, which correspond to different ways to group its elements of analysis and facts or measures. The relational implementation of the multidimensional data model is typically a star schema, or a snowflake schema [13, 14]. A star schema is a convention for organizing the data into dimension and fact tables. A snowflake schema is a variation of the star schema. Snowflaking is a form of dimensional modelling in which dimensions are stored in multiple related dimension tables.

Using these technologies we are able to publish scorecards implemented in a multidimensional data model using RDF and Linked Data technologies, obtaining a number of advantages as described by the W3C recommendation:

- The individual observations, and groups of observations, become (Web) addressable. This allows publishers and third parties to annotate and link to this data.
- Statistical data can be combined across datasets.

The last version of this article was published online before print, January 13,2016, Journal of Information Science, 2016, © The Author(s), DOI: 10.1177/0165551515324353,

http://jis.sagepub.com/content/early/2016/01/13/0165551515624353.abstract

- Publishing scorecards as Linked Data offers a flexible, non-proprietary, machine readable means of publication.
- It enables reuse of standardized tools and components.

For our work, some existing vocabularies and ontologies are used, such as:

- RDF data cube vocabulary[7] is a standard to publish multi-dimensional data, on the Web in such a way that it can be linked to related data sets and concepts. The current version of RDF vocabulary does not enable the aggregation of data from different granularity level along a dimension hierarchy. This vocabulary defines:
    - datasets, representing the container of some data;
    - dimensions, meaning some analysis criteria (for example a time period, location, etc.);
    - measures, representing a piece of data (e.g. a cell in a table), a KPI; and,
    - attributes, expressing characteristics of dimensions.
- Dublin Core[8] is a set of terms that is used to describe web resources as well as physical resources. Dublin Core Metadata may be used to provide interoperability in semantic web implementations. Some terms of this vocabulary are: dc:identifier, dc:title, dc:creator, dc:subject, etc.
- BIBO[9] (The Bibliographic Ontology) provides concepts and properties for describing bibliographic resources and relations on the semantic web using RDF. Terms of this ontology are: academic article, book, proceedings, object properties such as dc:title, dc:creator, rdf:about, and data properties such as bibo:edition, bibo:issue, bibo:volume, etc.
- FOAF[10] (Friend of a Friend) is an ontology describing persons, their activities and relations to other people and objects in RDF format. Some terms in FOAF vocabulary are: foaf:name, foaf:homepage, foaf:person, foaf:familyName, etc.
- ORG[11] (Organization) is the ontology for describing organizations, roles and organizational activities. Some terms from this ontology are: org:organization, org:agent, org:event, org:site, etc.
- SKOS[12] (Simple Knowledge Organization System) is a standard for sharing and linking concepts and concept schemes. Some terms from this ontology are: skos:concept, skos:collection, skos:semanticRelation, skos:mappingRelation, skos:closeMatch, skos:member, skos:topConceptOf, etc.
- VoID[13] (Vocabulary of Interlinked Datasets) allows express metadata about RDF datasets. VoiD covers four areas of metadata:
    - General metadata follows the Dublin Core model. Examples of terms are dcterms:title for the name of the dataset, and dcterms:license, to point to the license under which a dataset has been published.
    - Access metadata describes how RDF data can be accessed using various protocols. An example of access metadata is void:sparqlEndpoint.
    - Structural metadata describes the structure and schema of datasets and is useful for tasks such as querying and data integration. VoID also provides a number of properties for expressing numeric statistics about a dataset, such as the number of RDF triples it contains, or the number of entities it describes.
    - Linksets metadata describes links between datasets, it is helpful for understanding how multiple datasets are related and can be used together. An example of this kind of metadata is void:Linkset.

3. Evaluation of the Use of Open Access Journals

The evaluation approaches, methods, and criteria vary among the existing digital libraries (DL) evaluation studies [15, 16, 17, 18]. A DL is a collection of information stored in digital formats and accessible by computers [19]. OAJ are a type of specialized DL. The majority of the studies adopt Information Retrieval (IR) evaluation approaches at a restricted level (either at the system or the user level) while employing traditional criteria, such as precision, search time, error rate, etc. Very few evaluate the benefits of an OAJ on the user. Furthermore, there are few metrics devised specifically for this goal interlinked with external information. Due to this reasons, scorecards are an ideal candidate for covering these deficiencies.

The last version of this article was published online before print, January 13,2016, Journal of Information Science, 2016, © The Author(s), DOI: 10.1177/0165551515324353,

http://jis.sagepub.com/content/early/2016/01/13/0165551515624353.abstract

3.1. Scorecards

A scorecard is a tool to monitor progress toward a corporate goal in a business. The Balanced Scorecard is one of the best well-known corporate scorecards; it is used to help organizations to align them with their strategic objectives [20]. The overall strategic goals are broken down into a series of objectives that enable the organization to meet its strategic goals. Each of these objectives is associated with one or more KPIs, so progress towards each objective can be measured. KPIs are business metrics used to evaluate factors that are crucial to the success of an organization. In order to use KPIs, measures about actual value, target value and variance should be defined.

Table 2 shows a fragment of an OAJ scorecard. In this example, the OAJ has the following goal: Make self-diagnosis, an objective is to monitor trends over time. This goal requires monitoring performance against the targets defining KPIs such as the number of users and managing it through a scorecard. In the example the managers expect to increase the actual number of visits per month in a year up to 10.000.

Table 2. OAJ Scorecard Fragment.

Goal	Objectives	KPI	Actual value	Target value
Make self-diagnosis.	To monitor trends over time.	number of users/month	1,000 users/month	10,000 users/month

3.2. Scorecards and OAJ

Performance metrics and indicators should be related to institutional and OAJ mission and objectives [20]. But, analysing a random sample of OJS from DOAJ[14] (Directory of Open Access Journals), few of them publish their vision, mission, strategic objectives, or statistics.

The ISO 2789:2013 [7] standard defines statistics for "evaluation and comparison of libraries as well as for promoting, marketing and advocating the value that libraries provide for their population and for society". The objectives of the library statistics defined in the ISO 2789:2013 standard are summarized as follows:

- to monitor operating results against standards and data of similar organizations;
- to monitor trends over time;
- to provide a base for planning, decision making, improving service quality, and feedback of the results;
- to inform national and regional organizations in their support, funding and monitoring roles, and,
- to demonstrate the value of library services obtained by users, including the potential value to users in future generations.

For our work, we have developed a scorecard to:

- monitor use trends over time,
- make self-diagnosis, and,
- use the results in marketing.

Another related standard used is ISO 11620:2014 [21]. This standard specifies the requirements of a performance indicator for libraries and establishes a set of indicators to be used by libraries of all types. This international standard offers accepted, tested, and publicly accessible methodologies and approaches to measure a range of library services. Performance indicators can be used for comparing over time within the same library.

A primary purpose of using library performance indicators is self-diagnosis, including comparisons within the same library in several years [22]. We focus our study mainly on this requirement using Linked Data technologies to allow future analysis based on interlinked indicators.

The last version of this article was published online before print, January 13,2016, Journal of Information Science, 2016, © The Author(s), DOI: 10.1177/0165551515324353,

http://jis.sagepub.com/content/early/2016/01/13/0165551515624353.abstract

The proposed model can be used as a strategic scorecard which can also be navigated. We have used a subset of indicators of the ISO 2789:2013 and ISO 11620:2014 standard, for the use of electronic documents, based on interviews with librarians, local authorities and the data that was possible to retrieval from the OJS records.

The selected indicators are:

I1: number of virtual visits, (count the number of virtual visits on the library website, regardless of the number of pages or elements viewed, during the reporting period),
I2: number of rejected accesses (count the total number of unsuccessful requests of a licensed electronic services provided by the library by exceeding the simultaneous user limit),
I3: number of downloads (total number of successful content unit downloads requested from a library-provided online service),
I4: % external users (% of the library's total access from countries different to the country library),
I5: % of documents not used (% of documents not accessed),
I6: user satisfaction (The average rating by users of the library services),
I7: number of digital documents stored, and,
I8: number of digital documents added .

Along with these indicators extracted from the standard, we have included several dimensions of analysis that help in aggregating or disaggregating the information at hand:

D1: time (analysis time),
D2: article (published article),
D3: author (article author),
D4: geographic location (visiting geographic location),
D5: keyword (keywords defined in the articles), and,
D6: objective (OAJ strategic objectives).

The indicators I1-I5 could be analyzed for all de dimensions. The indicators I6-I8 could be analyzed for all the dimensions except for D4 (geographic location) and D2 (article). A monthly granularity is defined for all the measures.

4. Linked Data Publication Process for a Scorecard

In order to publish and feed a scorecard from an OAJ data mart transformed into RDF format we propose five main steps executed interactively as shown in Figure 5.

In the following sections 4.1 to 4.5, we describe each step in the proposed process from the data source identification and analysis to the publishing in a SPARQL platform (SPARQL end point).

4.1. Data Source Analysis

In this initial step, we analyse the information provided by the OAJ data source that could be useful for the proposed scorecard. This data source had the information about publications, which we needed to link with other datasets to give us better knowledge about the use of publications. First, we represented the OAJ data source in the form of a multi-dimensional data model, comprised of three basic components: dimensions, measures, and attributes. This allowed us to approach the data source as a data mart, a subset of the target data warehouse for OAJ evaluation.

Data marts are usually oriented to specific business topics (the topic in this case would be publications), and they allow us to build specialized scorecards for each area. The data mart is implemented in a multi-dimensional data model. The relational representation of the resulting multi-dimensional data model is a star schema or a snowflake schema. A star schema presents the data into dimension tables and fact tables. The snowflake schema is a type of star schema in which the dimension tables are partly or fully normalized. In Figure 6 we present a snowflake schema corresponding to the OAJ data mart. The fact table contains the KPIs or measures and the dimension tables the criteria of analysis (time, article, objective, geographical-location, and author). Dependency relationships are designed from the dimensions to the fact table. Dependency is a directed relationship to show that some elements depend on other model elements.

The last version of this article was published online before print, January 13,2016, Journal of Information Science, 2016, © The Author(s), DOI: 10.1177/0165551515324353,

http://jis.sagepub.com/content/early/2016/01/13/0165551515624353.abstract

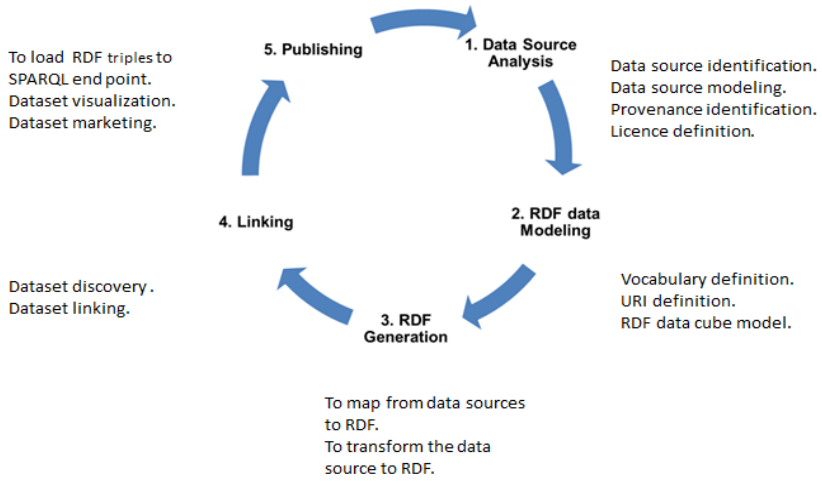

Figure 5. Linked Data publication process.

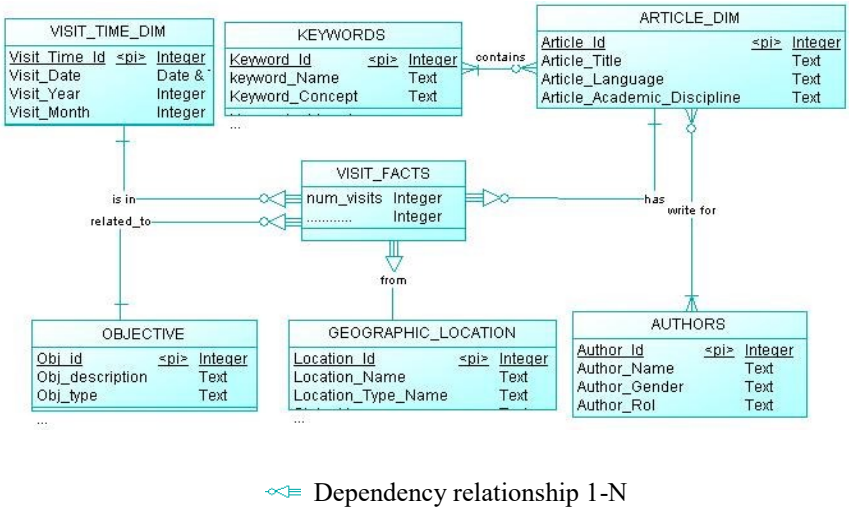

Figure 6. OAJ Visits Snowflake Schema.

This data linked to other datasets will give us better knowledge about similar subjects, the authors who work in them, and the objectives accomplished related to national goals. However, in order to be able to link this data, we need to transform it into RDF.

4.2. RDF Data Modeling

The goal of this step is to design and implement the vocabularies for describing the datasets in RDF. The most important recommendation from several studies [23, 24] is to reuse available vocabularies as much as possible to develop new ontologies [25]. To this aim, we use the controlled vocabularies and ontologies described in section II for modelling statistical datasets in RDF such as RDF data cube vocabulary, BIBO, Dublin Core, FOAF, ORG, SKOS, and VOID.

The last version of this article was published online before print, January 13,2016, Journal of Information Science, 2016, © The Author(s), DOI: 10.1177/0165551515324353,

http://jis.sagepub.com/content/early/2016/01/13/0165551515624353.abstract

The reduced RDF data cube model obtained as a result of this step is presented in Figure 7. In this RDF model, each concept is mapped with the corresponding concept of the model, such as dimension, measure, code list, etc. In RDF, each resource is identified by a URI. URI enables interaction with the Web using specific protocols. The URI structure for our proposal was defined by:

- Datasets are identified by: {base_URI}/}/dc/cube_name/dataset/{datasetName}. Example the ojsvisits dataset is represented by: http://opendata.epn.edu.ec/dc/ojs/dataset/ojsvisits. Dataset is a collection of statistical data corresponding to a defined structure.
- Data structure definition which defines the structure of one dataset referencing to a set of component specifications. It defines the dimensions, attributes and measures. It is identified by: {base_URI}/dc/cube_name/dsd/ {dataStructureName}. Example: http://opendat.epn.edu.ec/dc/ojs/dsd/dsd-ojs.
- The dataset component which stands for dimensions, measures and attributes represented as RDF properties in the Data Cube vocabulary, is specified by: {base_URI}/dc/ cube_name/prop /{dimension_name| measure_name}.For example the article dimension is represented by : http://opendata.epn.edu.ec/dc/ojs/prop/ article.
- Concepts and their values reused across multiple datasets are identified by: {base_URI}/concept/ {conceptName} and {base_URI}/concept/{conceptName}/{value}. Example: http://opendata.epn.edu.ec/concept/physics.

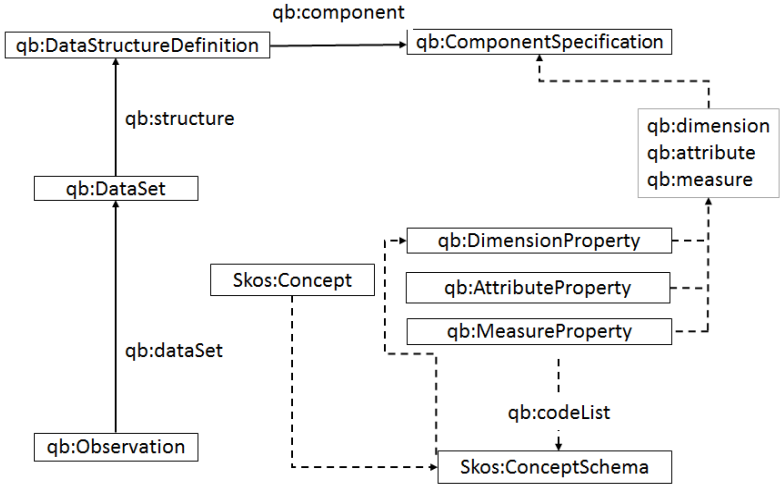

Figure 7. Outline of the RDF Data Cube vocabulary.

4.3. RDF Generation

The goal of this activity is to define a method and technologies to transform the source data into RDF and produce a set of mappings from the data sources to RDF. For the case study we used Open Refine[15] tool to perform the transformation from the multi-dimensional model stored in a relational database to a RDF data cube vocabulary.

Mappings were defined from the multidimensional database to RDF Data Cube elements, e.g., dimensions as qb:Dimension Property, measures as qb:Measure Property or attributes as qb:Attribute Property, the identification of the data (observations) as qb:Observation instances. Concepts within the datasets may be mapped with other concepts and code lists (controlled vocabularies) providing compatibility and interoperability. The mappings are used to create the dataset's structure, the dataset itself and the observations, using the appropriate URI Scheme for each type of resource [26, 27]. The code lists that are used to give a value to each of the components are also defined using SKOS vocabulary. The data are then exported as RDF in a RDF compliant serialization, such as RDF/XML as shown in Figure 8.

The last version of this article was published online before print, January 13,2016, Journal of Information Science, 2016, © The Author(s), DOI: 10.1177/0165551515324353,

http://jis.sagepub.com/content/early/2016/01/13/0165551515624353.abstract

4.4. Interlinking

The objective of this step is to improve the connectivity to external datasets enabling other applications to discover additional data sources. For this task we perform two steps: discovery, and linking.

```
<rdf:Description rdf:about="http://opendata.epn.edu.ec/dc/ojs/dataset/ojsvisits">
        <rdf:type rdf:resource="http://purl.org/linked-data/cube#DataSet"/>
        <rdfs:comment xml:lang="en">EPN Journal Visits (1/10/2015-31/10/2015)</rdfs:comment>
</rdf:Description>
<rdf:Description rdf:about="http://opendata.epn.edu.ec/dc/ojs /prop/article">
        <rdf:type rdf:resource="http://purl.org/linked-data/cube#ComponentSpecification"/>
        <qb:dimension rdf:resource="http://opendata.epn.edu.ec/dc/ojs/prop/article"/>
        <rdfs:label xml:lang="en">Article</rdfs:label>
</rdf:Description>
<rdf:Description rdf:about="http://opendata.epn.edu.ec/dc/ojs/dccs/ojsvisitsmeasure">
        <rdf:type rdf:resource="http://purl.org/linked-data/cube#ComponentSpecification"/>
        <qb:measure rdf:resource="http://opendata.epn.edu.ec/dc/ojs /prop/ojsvisitsmeasure"/>
        <rdfs:label xml:lang="en">Sessions</rdfs:label>
</rdf:Description>
```

Figure 8. Partial result of RDF/XML code generated.

Discovery comprises finding new target datasets. For this step we used the website "the Datahub"[16]. We found the DOAJ directory and several open linked datasets from scientific journals. Moreover we found statistics from several countries related to business, organizations and research topics. We will focus the analysis in the most visited articles looking for linking to similar topics in datasets like Dbpedia to increase the information about authors, research networks, organizations sponsoring similar works, research articles in similar topics, etc.

Linking allows us to relate external sources for additional information. For this step we used the open source software Silk[17] to find relations between data items in our datasets and the external datasets generating the corresponding RDF links that were stored in a separated dataset. This data will help us to develop new interrelated KPIs for example if we have number of visits by country extending the information from another dataset with number of students by country we could have number of visits/student by country. In addition we could link the keywords from the articles to keywords from research funding institutions to have information about visits from articles by funding institutions and so on.

4.5. Publishing

The goal of this activity is to make RDF datasets available on the Web to users, following the Linked Data principles. For this activity, we need a RDF server, usually in the form of a SPARQL endpoint. In our case the generated triples were loaded into a SPARQL endpoint (a conformant SPARQL protocol service) based on OpenLink Virtuoso[18] which is a database engine that combines: the functionality of RDBMS, virtual databases, RDF triple stores, XML store, web application server and file servers. On top of OpenLink Virtuoso, Cubeviz[19] and Ontowiki[20] component is used as a Linked Data interface to datasets complying to the RDF Data Cube vocabulary[28]. Datasets may be further "announced" to the public, to be more discoverable, by publishing the data to international or national open data portals. Figure 9 shows a view of the SPARQL endpoint with a partial result of the query on the OJS visits data cube, giving the number of visits by subject and by article. It will be possible in the future to annotate the relationships on the Web and to add more links increasing the knowledge and the possibility to get more complex queries.

In Figure 9, the y axis presents name of published articles in Spanish. The x axis contains the number of visits. Table 3 presents the translation from Spanish to English of the labels. In this graphic we can see the measure number of visits and two dimensions of analysis disciplines and articles

The last version of this article was published online before print, January 13,2016, Journal of Information Science, 2016, © The Author(s), DOI: 10.1177/0165551515324353,

http://jis.sagepub.com/content/early/2016/01/13/0165551515624353.abstract

Table 3: Translation of axis values in Figure 9.

Spanish	English
Aplicaciones de Procesamiento de Lenguaje Natural.	Applications of Natural Language Processing.
Estudio de la generación de hidrocarburos marcadores del proceso de irradiación en carne de cerdo.	Study of hydrocarbon generation markers of the irradiation process in pork.
Desarrollo de Modelos Digitales para la Dosimetría de Cobalto-60 de la Escuela Politécnica Nacional.	Development of Digital Models for Dosimetry Cobalt- 60 of the National Polytechnic School.
Evaluación experimental del problema de flujo no divisible de costo mínimo con única fuente mediante la aplicación de algoritmos genéticos.	Experimental evaluation of the problem of not divisible flow with minimal cost by applying genetic algorithms.
Algoritmos.	Algorithms.
Procesamiento de lenguaje natural.	Natural Language Processing.
Radiaciones.	Radiation.

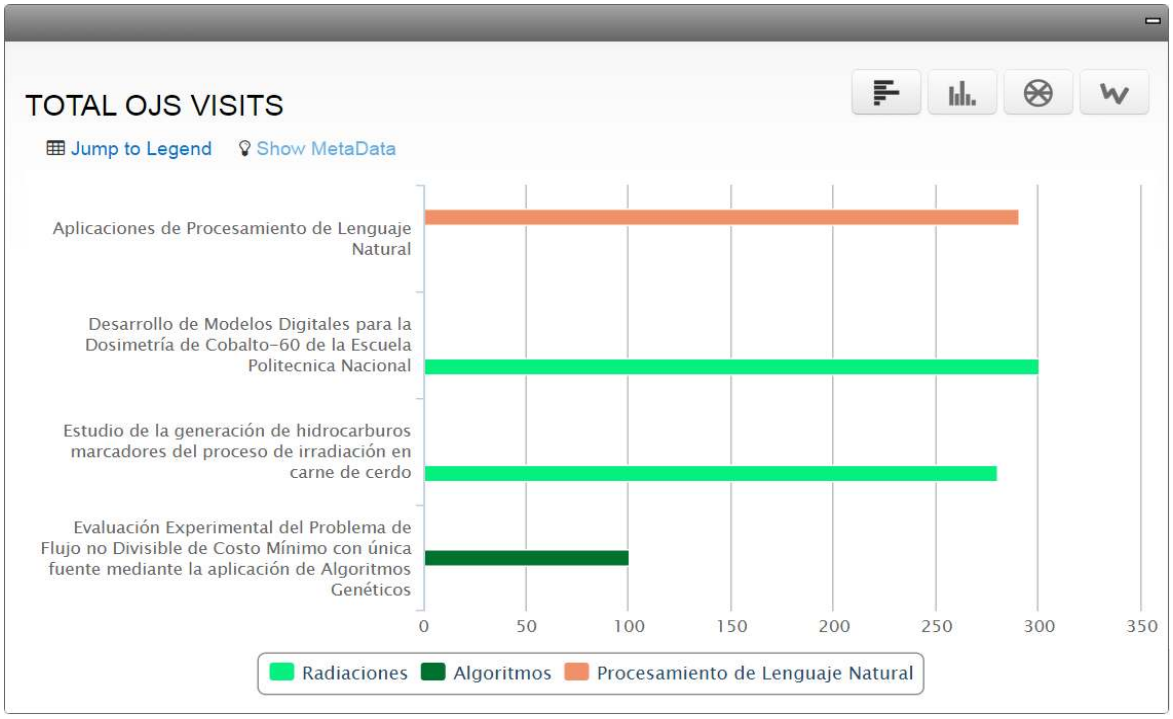

Figure 9. Query result example on the OJS visits data cube**.**

5. Technical Architecture

The architecture used in this proposal is shown in the Figure 10; Spoon software was used to extract metadata from OAJ, Open Refine software was used to transform data from the snowflake schema to RDF triples in RDF data cube vocabulary. The generated triples were stored in Open Link Virtuoso software and visualized using CubeViz software.

The users could also access to the RDF data using SPARQL language from virtuoso software. Silk software was used to discover related datasets for linking the RDF generated triples.

The last version of this article was published online before print, January 13,2016, Journal of Information Science, 2016, © The Author(s), DOI: 10.1177/0165551515324353,

http://jis.sagepub.com/content/early/2016/01/13/0165551515624353.abstract

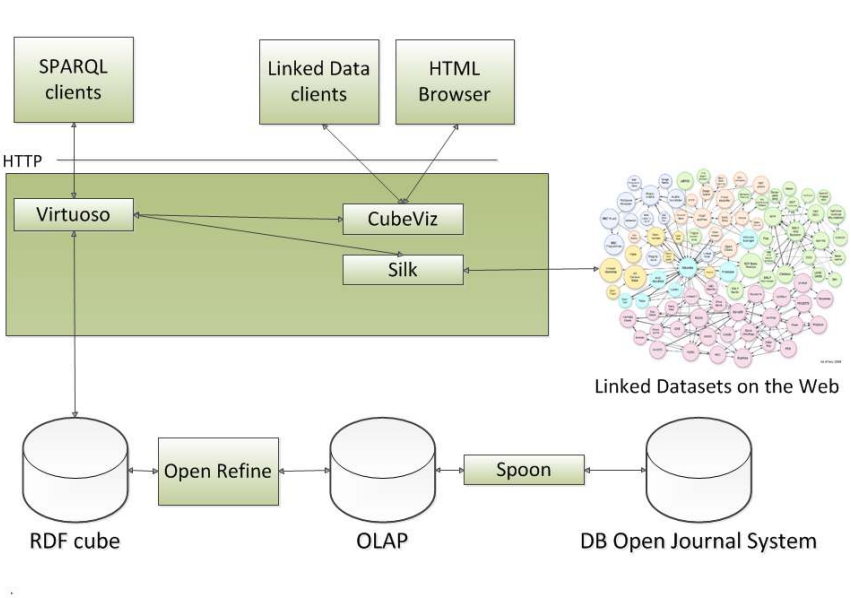

Figure 10. Architecture for Scorecard RDF Publishing.

6. Conclusions and Future Work

In this paper, we have described a process for evaluating the use of scientific data from Open Access Journals on the Web using scorecards and the principles of Linked Data. The process is based on best practices and recommendations from several studies, adding tasks and activities considered important during the project. The process begins with the scorecard development, the transformation into a multidimensional model and afterwards into RDF using the RDF data cube vocabulary. For publishing the RDF multidimensional model we used OpenLink Virtuoso, Onto-wiki and CubeViz applications. The Open Refine software was applied for the RDF generation process. In order to get better KPIs (Key Performance Indicators), the proposal also allows us to reuse existing and published information into RDF format. The traditional evaluation methods such as the proposal in Project COUNTER or the standards ISO 2789:2013 and 11620:2014 do not give the possibility of automatic linking of indicators and analysis features to external data. The power of linking measures with Linked Data goes far from hyperlinks, giving the possibility to annotate and reference statistical data and the nature of the relationships on the Web. In addition, it is possible to add dynamically more links to new resources. By providing context to the connection, it creates knowledge, because the link itself is knowledge. The proposed model can help us find new things inferred from the stored triples. As a result, the developed process fulfilled the requirements of the study.

In the future, we plan to develop a user registration interface, to be accessed before downloading the articles, in order to get more data for analyzing and comparing search history data. Moreover, we will design metrics to evaluate the performance of the proposed process for the development of new scorecards oriented to other strategic objectives. Furthermore, we will develop and look up for related open linked dataset catalogues to link projects results, enhancing the associated information and creating new interlinked KPIs. Finally, we are planning to develop a recommender system linking information to datasets from OAJs.

Notes

1. Revista Politécnica: http://www.revistapolitecnica.epn.edu.ec/. Revista Politécnica is a scientific journal from National Polytechnic School.

The last version of this article was published online before print, January 13,2016, Journal of Information Science, 2016, © The Author(s), DOI: 10.1177/0165551515324353,

http://jis.sagepub.com/content/early/2016/01/13/0165551515624353.abstract

2. W3C. Cool URIs for the Semantic Web, http://www.w3.org/TR/2008/NOTE-cooluris-20081203/.
3. W3C. RDF Resource Description Framework. http://www.w3.org/RDF/.
4. W3C. RDF/XML Syntax Specification, www.w3.org/TR/REC-rdf-syntax/ 2004.
5. W3C. RDF vocabulary description language 1.0: RDF Schema Recommendation, www.w3.org/TR/2004/REC-rdf-schema-20040210.
6. W3C. OWL 2 Web Ontology Language Document Overview (Second Edition). http://www.w3.org/TR/owl2-overview/.
7. RDF data cube vocabulary: http://www.w3.org/TR/vocab-data-cube/.
8. Dublin Core Metadata Element Set, version1.1:
9. The Bibliographic Ontology: http://bibliontology.com/.
10. The Friend of a Friend (FOAF) project: http://www.foaf-project.org/.
11. The Organization Ontology (ORG): http://www.w3.org/TR/vocab-org/.
12. Simple Knowledge Organization System (SKOS): http://www.w3.org/2004/02/skos/.
13. Vocabulary of Interlinked Datasets (Void): http://www.w3.org/TR/void/.
14. DOAJ: http://www.doaj.org. DOAJ is an online directory that indexes open access peer-reviewed journals.
15. Open Refine: http://openrefine.org/.
16. Datahub: http://datahub.io/. Datahub is a free data management platform used to publish RDF datasets.
17. Silk: http://wifo5-03.informatik.uni-mannheim.de/bizer/silk/.
18. Virtuoso Universal Server: http://virtuoso.openlinksw.com/.
19. CubeViz: http://cubeviz.aksw.org/.
20. Ontowiki: http://aksw.org/Projects/OntoWiki.html

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

References

[1] Harnad S. Open access scientometrics and the UK Research Assessment Exercise. Scientometrics 2009; 79(1): 147-156.
[2] Suber P. Open Access, London: The MIT Press Essential Knowledge Series, 2012.
[3] The university of Shefield. The University Library: Open Access Key Concepts, www.sheffield.ac.uk/library/openaccess/concepts#JournalTypes (2012, accessed 10 July 2015).
[4] Brian D and Willinsky E. A Survey of Scholarly Journals Using Open Journal Systems. Scholarly and Research Communication 2010; 1(2): 1-22.
[5] Hallo M, Luján-Mora S and Maté A, and Trujillo J. Current state of Linked Data in digital libraries. Journal of Information Science, Epub ahead of print 21 July 2015. DOI: 0165551515594729.
[6] Poll R and Payne P. Impact measures for libraries and information services. Library Hi Tech 2006; 4(4): 547-562.
[7] ISO 2789:2013. Information and documentation-International library statistics.
[8] Bizer C, Heath T, Idehen K, and Berners-Lee T. Linked Data on the Web. In Proceedings of the 17th international conference on World Wide Web, 2008, pp. 1265-1266.
[9] Banker R, Chang H, and Pizzini M. The balanced scorecard: Judgmental effects of performance measures linked to strategy. The Accounting Review 2004; 79(1): 1-23.
[10] Ermilov I. et al. Linked open data statistics: Collection and exploitation. Communications in Computer and Information 2013; 394: 242-249.
[11] Sánchez D, Batet M, Isern D, and Valls A. Ontology-based semantic similarity: A new feature-based approach. Expert Systems with Applications 2012; 39(9): 7718-7728.
[12] Setijono D, and Dahlgaard J. Customer value as a key performance indicator (KPI) and a key improvement indicator (KII). Measuring Business Excellence 2007; 11(2): 44-61.
[13] Luján-Mora S, Trujillo J, and Song I. Multidimensional Modeling with UML Package Diagrams. In: Lecture Notes in Computer Science 2503. Proceedings of the 21st International Conference on Conceptual Modeling (ER 2002) 2002, pp. 199-213.
[14] Luján-Mora S, Trujillo J and Song I. Extending the UML for Multidimensional Modelling. In: Lecture Notes in Computer Science 2460. Proceedings of the 5th International Conference on the Unified Modeling Language (UML 2002) 2002, pp. 290-304.
[15] Reeves T, Apedoe X, and Woo Y. Evaluating digital libraries: A user friendly guide, University Corporation for Atmospheric Research.2005, www.dpc.ucar.edu/projects/evalbook/EvaluatingDigitalLibraries.pdf. (2005, accessed January 15, 2015).
[16] Ying Zhang Z. Developing a holistic model for digital library evaluation. Journal of the Association for Information Science and Technology 2010; 61(1): 88-110.

The last version of this article was published online before print, January 13,2016, Journal of Information Science, 2016, © The Author(s), DOI: 10.1177/0165551515324353,

http://jis.sagepub.com/content/early/2016/01/13/0165551515624353.abstract

[17] Klas C. P, Albrechtsen H, Fuhr N, Hansen P, Kapidakis S, Kovacs L and Jacob E. A logging scheme for comparative digital library evaluation. Research and Advanced Technology for Digital Libraries 2006; 267-278.
[18] Pinto L,Ochôa P, and Vinagre M. Integrated approach to the evaluation of digital libraries: an emerging strategy for managing resources, capabilities and results. Library statistics for the 21st century world 2009; 273-288.
[19] Heradio R., Fernández-Amorós D, Cabrerizo F, Herrera-Viedma E, A review of quality evaluation of digital libraries based on users' perceptions. Journal of Information Science 2012; 38(3): 269-283.
[20] Maté A, Trujillo J, and Mylopoulos J. Conceptualizing and Specifying Key Performance Indicators in Business Strategy Models. In: Proceedings of the 2012 Conference of the Center for Advanced Studies on Collaborative Research 2012, p. 102-115.
[21] ISO 11620:2014. Information and documentation-Library performance indicators.
[22] Melo L, and Pires C. Performance evaluation of academic libraries: implementation model. Paper presented at: The 17th Hellenic Conference of Academic Libraries, 2008 September 24-26; Ioanina: Greece, 2008.
[23] Pesch O. Implementing SUSHI and COUNTER: A Primer for Librarians: Edited by Oliver Pesch. The Serials Librarian 2015, 69(2): 107-125.
[24] Uschold M and Gruninger M. Ontologies: Principles, Methods and Applications, Knowledge Engineering Review 1996; 11(2): 93-126.
[25] Keith A, Cyganiak R, Hausenblas M, and Zhao J. Describing Linked Datasets, In: Proceedings of Linked Data on the Web Worshop (LDOW2009) 2009.
[26] Hallo M, Luján-Mora S, Trujillo J. Transforming Library Catalogs into Linked Data. In: Proceedings of the 7th International Conference of Education, Research and Innovation 2014, pp. 1845-1853.
[27] Candela G, Escobar P, Marco-Such M, Carrasco R. Transformation of a Library Catalogue into RDA Linked Open Data. Research and Advanced Technology for Digital Libraries 2015, pp. 321-325.
[28] Mader C,Martin M, and Stadler C. Facilitating the Exploration and Visualization of Linked Data. In: S. Auer et al. (ed.) Linked Open Data-Creating Knowledge Out of Interlinked Data, LNCS 8661, London: Springer, 2014, pp. 90-107.

The last version of this article was published online before print, January 13,2016, Journal of Information Science, 2016, © The Author(s), DOI: 10.1177/0165551515324353,

http://jis.sagepub.com/content/early/2016/01/13/0165551515624353.abstract