AGILE: GIScience Series, 3, 44, 2022. https://doi.org/10.5194/agile-giss-3-44-2022
Proceedings of the 25th AGILE Conference on Geographic Information Science, 2022.
Editors: E. Parseliunas, A. Mansourian, P. Partsinevelos, and J. Suziedelyte-Visockiene.
This contribution underwent peer review based on a full paper submission.

Updating and using the EO4GEO Body of Knowledge for (AI) concept annotation

Rob Lemmens[1], Florian Albrecht[2], Stefan Lang[2], Sven Casteleyn[3], Martyna Stelmaszczuk-Górska[5], Marc Olijslagers[4], Mariana Belgiu[1], Carlos Granell[3], Ellen-Wien Augustijn[1], Carsten Pathe[5], Eva-Maria Missoni-Steinbacher[2], Aida Monfort Muriach[3]

[1]Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, The Netherlands
[2]Department of Geoinformatics, University of Salzburg, Salzburg, Austria
[3]Institute of New Imaging Technologies (INIT), Universitat Jaume I (UJI), Spain
[4]Spatial Applications Division Leuven, KU Leuven, Belgium
[5]Department of Earth Observation, Friedrich-Schiller-University, Jena, Germany

Correspondence: Rob Lemmens (r.l.g.lemmens@utwente.nl)

Abstract. The EO4GEO Body of Knowledge (BoK) serves as a vocabulary for the domain of geoinformation and earth observation, supporting the annotation of online resources. This paper presents how the BoK is designed, maintained and improved. We discuss how the BoK content can be extended, using the example of integrating artificial intelligence (AI) concepts and show how annotation is done by adding persistent concept identifiers in the metadata of training materials. This platform allows us to share online information with clarified semantics. A prolonged use necessitates the incentivisation of an active expert community and a further adoption of infrastructure standards.

Keywords. Body of Knowledge, concept model, concept annotation, artificial intelligence

1 Introduction

Information integration plays an important role in our ever increasing digital data consumption. Within the integration, it is often key to make explicit what are the common semantics on which the integration is based. This can be done by specifying definitions of common concepts and the relationships between them, in a *Body of Knowledge* (BoK). DiBiase et al. (2007) defines a Body of Knowledge as "a comprehensive inventory of the intellectual content that defines a field". The EO4GEO project started in January 2018 with reviewing and extending the GIS&T BoK developed by the GI-N2K project (Vandenbroucke and Vancauwenberghe, 2016). This BoK contained concepts related to geoinformation (GI) but required an expansion of its rudimentary content on Earth Observation (EO). EO4GEO is set up as an Erasmus+ Sector Skills Alliance and aims to bridge the gap between supply and demand of education and training in the EO and GI sectors. This is done by building a BoK together with a number of tools for generating online resources that allow stakeholders to build and find curricula, occupational profiles, job offers, learning material etc. The BoK provides a common vocabulary that the tools apply to annotate their provided content for improved search and comparison of online resources.

2 BoK updating and versioning

2.1 Creating and maintaining concepts

In the BoK, the concepts are the building blocks for representing the EO*GI knowledge domain. They are created and maintained in 7 working groups covering 14 top level domains. Each concept in the BoK contains a set of elements as below.

Code - Unique identifier assigned by working group leader, e.g., [AM6], [PP2-3-1].

Name - Identification of the concept in 1-5 words.

Description – Plain narrative text (max. 500 words without pictures or tables), formulated as an abstract about the concept. It should be self-contained and accessible to a wider audience.

Skills – Representation of the ability to apply knowledge and use know-how to complete tasks and solve problems in a certain occupational profile. For the design of curricula and learning material, skills act as learning objectives. Skills are formulated as verb + a concept relevant statement, using active voice. The verbs used for skills descriptions are action words borrowed from Bloom's taxonomy (Bloom et al., 1956) that was used in a revised version for the purpose of this BoK (Hofer et al., 2020). The taxonomy provides six levels of increasing cognitive skill, i.e. (1) remember, (2) understand, (3) apply, (4) analyse, (5) evaluate, and (6) create, where each category contains 15 to 25 verbs. For example, the level 2 "understand" includes verbs like cite, explain, interpret, and discuss. A skill example from the BoK concept 'Image classification' is 'Explain the advantages of object-based classification approaches over pixel-based approaches'. The full taxonomy is included in Hofer et al. (2020).

References – Authoritative sources used for the concept description and for further understanding of the concept.

Relationships between concepts – Outgoing and incoming relations to/from other concepts, typically Resource Description Framework (RDF)-triple style, with the current concept (subject) – relationship (predicate) – another concept (object) (see Ronzhin et al., 2018) with three types of relationships:

> *Subconcept* - a concept on a lower granularity level. A concept can be a subconcept of more than one other concept.
>
> *Prerequisite* - a concept needs to be known to understand the other.
>
> *Similarity* - a concept is almost the same as another.

Concept status – Indication of the status of the concept in the revision/definition process in the BoK. The status may be *deprecated, new, in progress, completed*, etc.

The BoK can be explored in two ways, i.e., through a concept map view (see Figure 2) used for editing the BoK and through a hierarchical view (see Figure 3), used for general exploration.

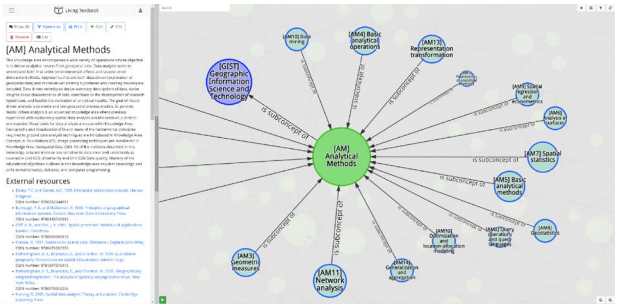

Figure 2. Part of the EO4GEO BoK in the Living Textbook[1]. The Living Textbook tool facilitates the creation of ontologies and visualises textual content and a concept map, side by side.

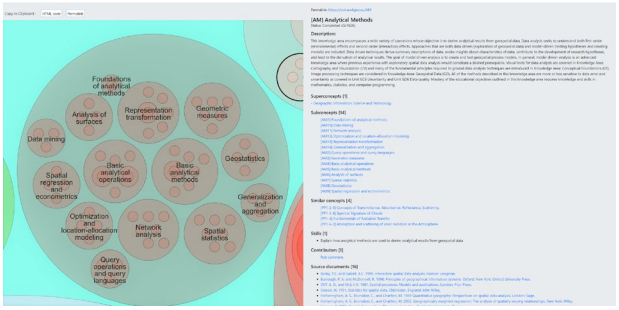

Figure 3. Part of the EO4GEO BoK in the BoK Visualisation and Search (BVS) tool[2]. BVS facilitates a synchronised graphical and textual view, advanced search, permalinks and BoK versioning.

Concepts are being maintained by the abovementioned working groups, through calls for experts, in which one or more concepts are targeted for an update. For missing topics, new concepts are created. The content of the concepts is either captured first in a text form or entered directly in the Living Textbook tool. Apart from such asynchronous updates, occasionally, events are used to update concept content collaboratively, such as during the EARSEL workshop on June 9th 2021 in Warsaw, Poland (virtual)[3]. During this workshop, participants contributed by generating descriptions of skills in the context of two concepts: the methodological concept *Image classification* and the application-oriented concept *Monitor urban areas*. This interactive discussion format worked well for skills because they are short and flexible to be attached to any concept of the BoK. The participating EO experts first identified new skills and afterwards discussed the quality of their descriptions. This resulted in easy-to-integrate high quality content for the BoK.

[1] http://www.eo4geo.eu/tools/living-textbook/

[2] http://www.eo4geo.eu/tools/bok-visualization-and-search/

[3] http://www.eo4geo.eu/earsel-eo-educationworkshop/

2.2 Releases

As a rule of thumb, a new official version of the BoK is released every three months. In parallel, a working version, the so-called *Sandbox* BoK, is continuously being updated. Before each release, the procedure is as follows:

i) announcement of the next BoK release cycle

ii) finalisation of the concepts in progress and approval by the EO4GEO Editorial Board, e.g., status of the concept is checked, ID code given

iii) copy of the current *Sandbox* BoK to an intermediate version that will be prepared for the new release

iv) final corrections and preparation of the *Master* version N +1 of the BoK, *Master* version N is archived

v) removal of the intermediate version of the BoK

vi) export of the *Master* version N+1 to the BoK platform at bok.eo4geo.eu.

The released version of *Master* BoK is visible in the Living Textbook tool for the public and can be commented on for future evolution. In the BoK platform, it is installed as the latest BoK version, made available through an API and used as a basis for the BoK Visualisation and Search and other BoK tools.

2.3 Persistent identifiers

To link or (re-)use the BoK, external sources and online applications need to refer to BoK concepts. To do so, permalinks are used to uniquely identify each BoK concept, e.g., https://bok.eo4geo.eu/AM for the *[AM] Analytical Methods* concept. To ensure backward compatibility, working groups carefully evolve concepts by updating their content or merging the content of two concepts into an existing one. Only truly new concepts are assigned a new identifier. Obsolete concepts are removed from the current version of the BoK, but they are still available as previous concepts in the BoK platform, so that links created in older external sources to those concepts do not break and an analysis on the evolvement of concepts is possible.

2.4 Link with other BoKs

The EO4GEO BoK is positioned as a reference for domain specific content, partly based on earlier initiatives GI-N2K and UCGIS GIS&T BoK (Wilson, 2014). Related content can be found in other vocabularies, such as the Cartographic BoK[4] and OGC Definitions Server[5]. Despite its focus on EO and GI, the EO4GEO BoK concepts contain information which overlaps with other domains, such as imaging, statistics and AI, and domains of more generic nature, for which other BoKs are in place, such as systems engineering[6] and data management[7]. Though many related domains disseminate their content, they often lack an interoperable exchange mechanism (Martirano et al., 2019) and therefore hinder an integral BoK approach through APIs or based on linked data. EO4GEO has made an effort to overcome this by offering a BoK platform, which offers API access and persistent identifiers (see Section 2.3) for annotations and referencing BoK concepts. New BoKs can be built for further specification of subdomains, or for neighbouring domains, by referring to and building on top of existing EO4GEO BoK concepts.

3 AI concepts in the BoK

The previous EO4GEO BoK versions contained only a few concepts related to AI, basically those that are affiliated with image understanding and computer vision in Earth Observation (EO). A necessary update that was performed recently is described below.

3.1 Integration procedure

In case of a larger update, existing concepts may be split or entirely new concepts are created. In all cases, existing relationships are checked and new relationships are created. Concerning AI, there were many concepts missing. Furthermore, AI content involves connections to multiple topics already in the BoK. To avoid the ad hoc adding of concepts to different content areas in the BoK we took a more centralised approach. First, a hierarchy of AI concepts was constructed, with 57 new concepts and 22 existing concepts. The top concept [AI] contains eight subconcepts: *Signal processing, Computer vision in EO, Computational linguistics, Automated reasoning, Machine learning, Hybrid AI, Intelligent software agents* and *AI algorithms*. The collection and grouping of the initial list of AI-related concepts was realised in a collaborative Google spreadsheet environment with several experts from the EO4GEO network contributing. The eight main pillars were established based on literature review and a selective choice of those AI concepts that have a direct connection to the EO*GI domain. Arguably there are more classifications possible, but this one was chosen as a start by the community of experts, and, driven

[4] https://bok.cartography.no/keywords-2/keywords/
[5] https://www.ogc.org/def-server
[6] https://www.sebokwiki.org/wiki/Guide_to_the_Systems_Engineering_Body_of_Knowledge_(SEBoK)
[7] https://www.dama.org/cpages/body-of-knowledge

by discussion, anticipated to converge to a commonly acceptable structure. After all, the BoK allows multiple-inheritance of concepts, and consequently, AI concepts can relate to other concepts, outside the abovementioned hierarchy.

Figure 5 shows some of the AI concepts that were created as a separate cluster.

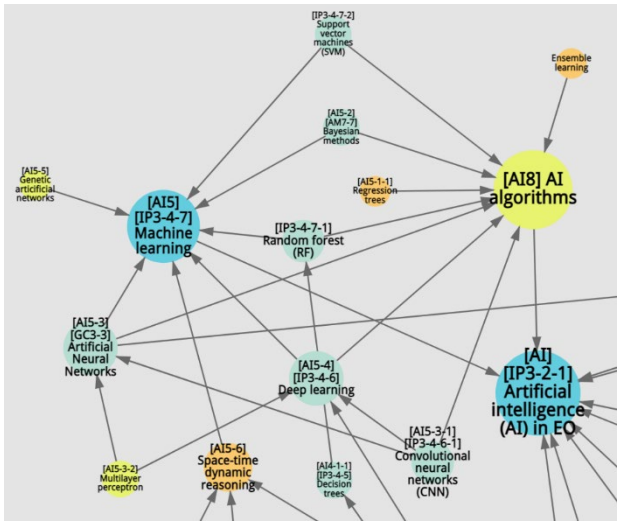

Figure 5. Part of the separate cluster of AI concepts before merging them into the actual BoK. The hierarchy is dominated by the AI codes (between square brackets); the original codes of existing concepts are held as additional codes. For the purpose of further integration, the concepts are marked with different colours: *Yellow*: New AI concepts; *Blue*: Existing concepts, which need content change; *Orange*: Existing concepts to be checked; *Green*: Existing concepts, which do not need a change of content.

79 AI concepts were then copied as a cluster to the existing BoK (900+ concepts), see Figure 6.

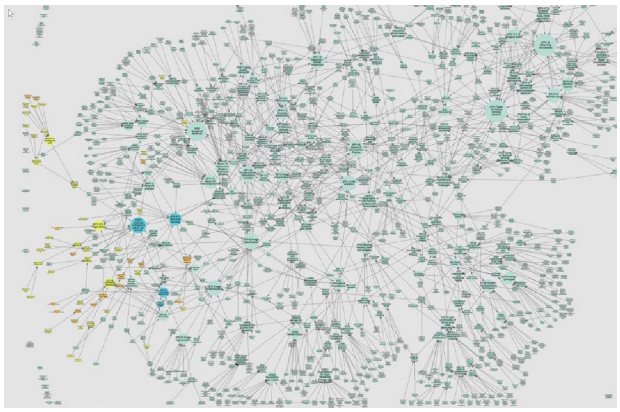

Figure 6. AI cluster integrated in existing BoK. The AI cluster appears on the left (the yellow, orange and blue coloured concepts; for colour codes, see the caption of Figure 5).

In the integration process, the AI concepts were all made subconcepts of existing concept domains, i.e., [GC] Geo computation, [AM] analytical methods and [IP] Image processing (see Figure 7).

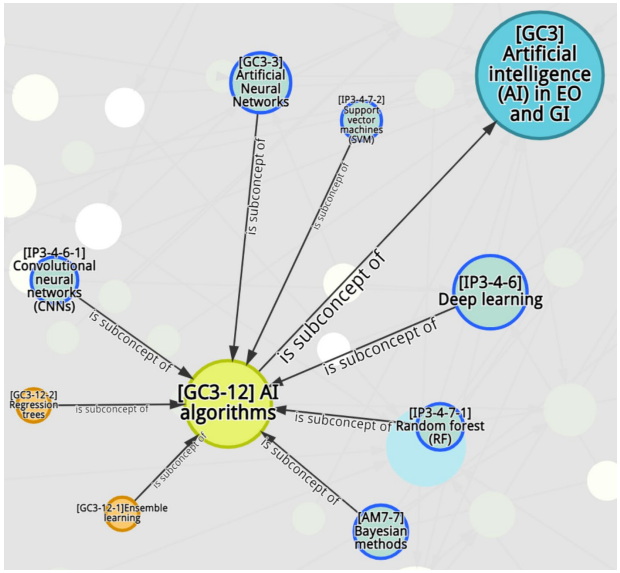

Figure 7. Part of the AI concept cluster after the merge into the existing BoK. Note: In this figure, the focus is on the *AI algorithms* concept and only direct subconcept relationships with the *AI algorithms* concept are shown. In the BoK, *Convolutional neural networks (CNNs)* is also a subconcept of *Deep learning*, but this is not shown here.

3.2 AI concept example

Random forest (Breiman, 2001) has become a popular machine learning method in remote sensing and other domains due to its high performance (Belgiu and Drăguţ, 2016). In the BoK, the *Random forest* concept (Albrecht et al. 2022) is a subclass of *Machine learning* and of *AI algorithms*. The *Decision trees* concept is set as a prerequisite to *Random forest*, which means that *Decision trees* are recommended to be known prior to the *Random forest* concept.

An example of a skill (Bloom level 2) for this concepts is: 'Explain the sensitivity of the Random Forests classifier to the number of decision trees and the number of variables used to split the tree nodes.'

4 Annotation with BoK concepts

As concepts in the BoK have at least a basic description and each concept has a persistent identifier, they can be used to clarify the semantics of information by annotating documents, web pages, job offers, webinars, etc. EO4GEO has built a number of annotation tools and created a repository of annotated resources that can be searched based on BoK concepts.

4.1 EO4GEO ecosystem of tools

Based on the BoK, an ecosystem of tools was developed within the EO4GEO project: BoK Visualisation and Search, Occupational Profile, Job Offer, Curriculum

Design, BoK Annotation and BoK Matching tool[8]. All tools use the BoK, available through the EO4GEO platform's API. For example, in the Curriculum Design Tool, content of courses is specified in terms of BoK concepts, while learning objectives are specified in terms of skills. Created resources are exportable as PDF files, which provide a pretty-printed human view, along with machine-readable (semantic) annotations embedded in the PDF's metadata, expressing information such as type of resource, title and the knowledge (BoK concepts) conveyed. These are described as RDF triples using the Dublin Core metadata element set - similar as in Figure 8. Finally, resources are comparable using the BoK Matching Tool, based primarily on their BoK annotations.

4.2 EO4GEO training material

During the project, EO4GEO has organised several courses, webinars and workshops. Much of the training material has been made available on the project website. To create the training material in a uniform look and feel, the reveal.js[9] platform was used. Reveal.js facilitates the creation of web-accessible presentations and incorporation of metadata into HTML code, allowing for the inclusion of BoK annotations. Figure 8 shows the metadata entered for one of the courses, in the form of RDFa annotations, including general information and all BoK concepts covered by this course. The BoK concepts are indicated by their code, e.g., *[IP3-5] Image segmentation* (third <link> line in Figure 8), each of which is represented by a persistent identifier, in this case https://bok.eo4geo.eu/IP3-5. The rendered HTML page with course information is shown in Figure 9. The actual slide presentation can be found at the project website[10].

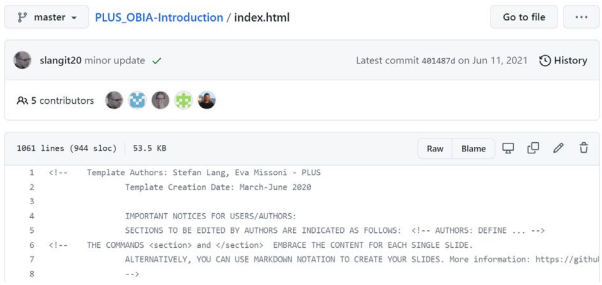

Figure 8. Excerpts of the BoK-based RDFa annotations entered in reveal.js for the introduction on Object-based image analysis, as part of the GEOBIA summer school in 2020.

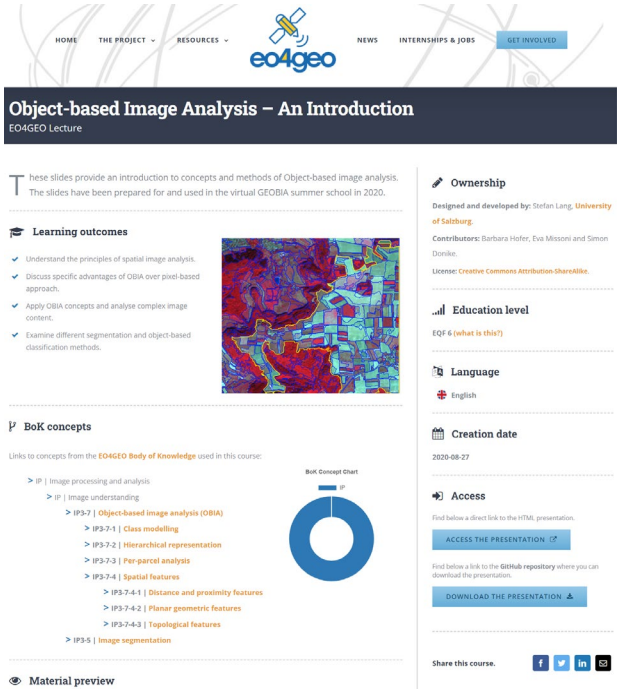

Figure 9. Presentation of meta-information about the introduction on Object-based image analysis, as part of the GEOBIA summer school in 2020. The covered BoK concepts are indicated by its hierarchy in an indented list.

4.3 Geo Course Hub

The Geo Course Hub project started in 2022 at the University of Twente to enable flexible course creation by modularising education into fine-granular shareable content, connected through learning paths. Course content, learning paths and learning objectives will be annotated with EO4GEO BoK concepts to allow searching and matching. Expectedly, new concepts are needed for application domains, such as water management and natural resource management. These will be held in an 'application' BoK that will be connected to the EO4GEO BoK. In addition, learning objectives will be formulated at all levels of Bloom's taxonomy. One of the limitations of the current EO4GEO BoK is that the

[8] http://www.eo4geo.eu/tools/
[9] https://revealjs.com/
[10] https://eo4geocourses.github.io/PLUS_OBIA-Introduction/

number of skills provided with each concept vary per concept.

4.4 Data and Software Availability

The EO4GEO BoK and its tools are publicly available at the EO4GEO website[11]. The BoK tools are licensed under GNU GPLv3.

5 Discussion and conclusion

The EO4GEO project has made an effort to create a vocabulary for EO and GI which is maintainable online and allows for annotating online resources. Although there are not many technical limitations, there are some hurdles to be taken to make the EO4GEO BoK a commonly accepted reference for new knowledge domains, such as (EO*GI-related) AI.

First, the BoK needs to keep a domain focus, to be able to manage its content within the EO*GI community. At the same time, links are needed from and to other domain vocabularies. This is certainly the case with AI, where many concepts are generally applicable in different domains of computer science. The use of generic vocabularies beyond Dublin Core, such as PROV-O[12], can be useful to manage the versioning of annotated sources. Linking vocabularies requires a proper technical infrastructure and well-established content management.

Second, for facilitating annotation of online resources, the concepts need to be findable and accessible over the Web. Besides the implementation with permalinks it would require additional standardised approaches such as those adopted in Linked Open Data (see Ronzhin et al. 2018). Schema.org can serve as a reference for reaching a wider audience.

Third, EO*GI is an innovative sector with several emerging and quickly evolving technical (r)evolutions. Maintaining a trends watch or following one or more existing trends watches is therefore an indispensable part of the future BoK update process. An active community of experts needs to keep the BoK up to date. EO4GEO has implemented a method for the inclusion of new concepts, split, merge and deprecate concepts and adapting relationships between concepts. Published articles in relevant journals serve as a basis for concept content. Motivating experts remains a challenge. Further work needs to be done on providing more incentives, such as lifting concept descriptions to micro publications.

References

Albrecht, F., Belgiu, M., Lang, S.: Random Forest - Concept in the EO4GEO Body of Knowledge, version 7.0. (in preparation), 2022. https://bok.eo4geo.eu/IP3-4-7-1

Belgiu, M., Drăguţ, L.: Random forest in remote sensing: A review of applications and future directions, ISPRS Journal of Photogrammetry and Remote Sensing, Volume 114, Pages 24-31, 2016.

Bloom, B.S., Krathwohl, D.R.: Taxonomy of Educational Objectives. The Classification of Educational Goals: Handbook 1: Cognitive Domain. Longmans, New York, 1956.

Breiman, L.: Random Forests. Machine Learning 45, 5–32. https://doi.org/10.1023/A:1010933404324, 2001.

Hofer, B., Casteleyn, S., Aguilar-Moreno, E., Missoni-Steinbacher, E., Albrecht, F., Lemmens, R., Lang, S., Albrecht, J., Stelmaszczuk-Górska, M., Vancauwenberghe, G., Monfort-Muriach, A.,: Complementing the European earth observation and geographic information body of knowledge with a business-oriented perspective. Transactions in GIS, 24, 587-601. https://doi.org/10.1111/tgis.12628, 2020.

Martirano, G., Bilotti, L., Vercillo, A., Linking the BoK for GI and EO to other BoKs. EO4GEO public report, Deliverable D 2.4. Available at http://www.eo4geo.eu/public-deliverables/, 2019.

Ronzhin, S., Folmer, E., Lemmens, R.: Technological Aspects of (Linked) Open Data. In: van Loenen, B., Vancauwenberghe, G., Crompvoets, J. (eds) Open Data Exposed. Information Technology and Law Series, vol 30. T.M.C. Asser Press, The Hague. https://doi.org/10.1007/978-94-6265-261-3_9, 2018.

Vandenbroucke, D., Vancauwenberghe, G.: Towards a New Body of Knowledge for Geographic Information Science and Technology. Micro, Macro & Mezzo Geoinformation, 2016 (6), 7-19. ISSN: 1857-9000, EISSN: 1857-9019, 2016.

Wilson, J. P.: Geographic Information Science & Technology: Body of Knowledge 2.0 Project (Final report), 2014. Retrieved from https://ucgis.memberclicks.net/assets/docs/ucgis_bok2_wilson_report_dec20 14.pdf

[11] http://www.eo4geo.eu/bok/

[12] https://www.w3.org/TR/prov-o/