Dept. of Civil, Environmental and Geomatic Engineering
University College London
Gower Street
London
WC1E 6BT
UK

26th May 2013

Re: Using Free and Open Source GIS to Automatically Create Standards-Based Spatial Metadata in Academia - First Investigations

Dear Editors

Thank you for your on-going assistance with the above paper. As requested, we have now incorporated the suggested revisions, details of which can be found below.

Kindly also extend my thanks to the reviewers for their very helpful and encouraging comments. We found the quality and relevance of the feedback extremely useful and feel that it has in turn helped to improve the quality of the paper.

Please do not hesitate to contact me should there be any further questions or suggestions, either at the address above or via the contact details provided below.

Best regards

Claire Ellul

E-mail: c.ellul@ucl.ac.uk

Comments from Organisers

Feedback	Comment
Your paper is just within the stated maximum word length (8000), but the large amounts of tables and figures makes the number of pages (26) so large that publication in the current form will be difficult, if not impossible. Please try to limit the number of pages to a maximum of 12-16, probably best done by excluding the large tables and some figures (you can make these available on the web and refer to those).	Agreed. Both long tables have now been shortened and full versions uploaded online for reference. A number of figures have also been removed. Length now reduced to 16 pages in total. The full tables are currently hosted at http://www.mapmalta.com – if the journal provides its own hosting for such information, kindly let me know.

Reviewer 1

Feedback	Comment
Authors should review sub-section 1, which currently reads not too well, also because it contains not appropriate references (e.g. quotes without a page reference, full sentence quotes without connection to previous statements). In the last paragraph of sub-section 1 it is not very clear what 'issues' authors address in the paper – this should be made more specific.	We feel that the issues referred to are clear from the previous paragraph – e.g. the complexity of metadata, the fact that it is decoupled from the data and the fact that it is often created at the end of a project.
Throughout the document authors seem to confuse INSPIRE with a standard – as they correctly state at the beginning of Section 2, INSPIRE is an infrastructure that uses standards for (among other things) medatada, more specifically INSPIRE uses guidelines for metadata creation based on EN ISO 19115 and EN ISO 19119.	Agreed. Text in Section 2 clarified to state that when we refer to INSPIRE metadata we are referring to ISO 19115.
Statement "As with any Spatial Data Infrastructure, metadata forms a core component of INSPIRE, and is based on ISO 19115 – this is not true: 1. Surely not any spatial data infrastructure is based on ISO 19115 (there are other metadata standards widely used in geocommunity – e.g. FGDC, Dublin Core)	Agreed. Text clarified to state that it is INSPIRE that links to ISO 19115, not ALL metadata.
Sub-sub-sections 2.2.1 – 2.2.4 do not add much to the paper and in my view should be reduced to an overview table displaying existing GIS software, its metadata capability and its limitations. This reduction will reduce the paper, which is quite lengthy at the moment.	Agreed. These sections reduced in scope.
Sub-section 3.2 – it seems that 'Identifier code' is simply an OID of the dataset as registered in the system, which is certainly different from URI	Agreed. Text clarified to explain that the OID will be used as part of a unique URI (project specific) for the dataset.

as proposed by the INSPIRE guidelines authors seem to follow. In my view this is insufficient and requires more reflection in the discussion, especially because authors mention future use of their system in web environment (e.g. WFS and CSW). In the future, how will URI be assigned to a data+metadata resource?	
Section 5 – in the footnote on p.16 authors say: "The data itself is also FOSS" – what is meant here? How can data be free and open source software (FOSS)?	Agreed. Sentence re-written.
Sub-sub-section 4.2.1 – bullet 3. "resource ID" is called "Identifier code" in Table 1 •	Agreed. Bullet changed for consistency.
What is the purpose of Figure 9 on p. 20?	Figure 9 helps to explain that the approach described in the paper automatically generates a catalog map of the datasets available to the user, that can be used by multiple GIS – i.e. is interoperable. However, figure removed due to space restrictions.
p. 20 - There is no sub-section 5.0 and neither sub-sub-section 5.0.1 so how can there be sub-sub-section 5.0.2? I think it is a good idea to explain automated metadata update, but 5.0.2 is very short and does not do the job.	Agreed. However, space does not permit a fuller explanation of the testing process and this sub section has now been removed.
Figure 10 is not legible and should be enlarged.	Agreed. This figure has now been removed due to space issues.
In section 6 authors state "…the functionality to maintain dataset and medatata synchronized is interoperable across multiple FOSS and non-FOSS GIS platforms" – this has not been tested and demonstrated; authors only show extension to the SPIT plugin in QGIS which takes shapefile format (non-FOSS GIS) and transforms it into a (FOSS) PostGIS/PostgreSQL table. It is useful, but it certainly does not demonstrate interoperability of the proposed solution across variety of platforms.	Agreed. Text clarified to explain that, due to the presence of triggers in the database, any data EDITS from any GIS will automatically result in updated metadata. It is the dataset/metadata synchronisation (rather than the initial metadata creation) that is interoperable.
It is not clear from the paper how do authors deal with shapefiles that have already metadata (e.g. as ISO 19139 compliant XML file). Is this ignored and new metadata is created?	A very good point – and one which we have not yet addressed in our approach. This comment has been added to the 'further work' section.
• p. 23 – "Web Catalog Service" should be "Catalog Service for the Web"	Agreed. Text updated.
References section needs serious revision – it contains spelling errors in the names of authors (e.g. Burrough) and many incomplete references.	Agreed. References revised. Some details not shown due to bibtex formatting.

Reviewer 2

Feedback	Comment
The title accurately reflects the content of the paper, except that it is not clear why the title and this work are restricted to an academic context. Refer also to the comment under 'Review' below. Abstract Similar to the title, the abstract accurately reflects the content of the paper, except that it is not clear why this work is restricted to an academic context. Refer also to the comment under 'Review' below.	Agreed - see below. A number of sentences have now been added to clarify why the academic context is relevant. The conclusion of the paper has also been updated to reflect that this work is also applicable elsewhere.
Keywords I would suggest to add 'automation' and/or 'metadata automation' to the list of keywords.	Agreed. Key words added.
Review Metadata generation and maintenance remain a challenge for which solutions need to be sought. This paper presents first investigations into a novel approach for metadata automation. The authors provide a logical justification for the research that refers to relevant literature. The paper is interesting and relevant to the FOSS4G2013 target audience, because open source tools are used. The paper is equally relevant to the wider geospatial community.	We thank the reviewer for their kind comments.
The stated objective of the paper is mostly met. The authors describe how they automated the creation of 18 of the 20 INSPIRE mandatory metadata elements. The data and metadata are tightly coupled in that they are stored in the same database, but it is not clear whether the workflow is tightly coupled. The authors refer to the 'tightly coupled' characteristic in two ways: tightly coupled in terms of storage (is the data and metadata integrated?), as well as tightly coupled in terms of workflow (is the metadata updated as part of the spatial editing workflow?). The approach described in the paper is definitely tightly coupled in terms of storage, but there is not enough information to evaluate whether the workflow is also tightly coupled. For example, when will the keywords be updated?	Agreed. The text has been clarified to state that the presence of the triggers means that keywords are automatically updated when the dataset is edited – resulting in tight coupling.
The claim that the approach described in the paper is interoperable needs to be better qualified in the abstract, introduction and conclusion.	Agreed (comment similar to reviewer 1 above). In particular, the conclusion now includes clarification as to which elements of the approach are interoperable to date, and

	where further work towards interoperability is required.
1. Introduction There is no overview of the paper in the introduction (to describe how different sections contribute to the stated objective) and the sections also do not have an introductory paragraph to explain how their content contributes to the paper's objectives. This leaves the author to read through individual sections in order to understand their contribution. Either the overview paragraph or introductory text for each section needs to be added.	Agreed. A short overview has been added to the introduction.
The definition for interoperability used in the ISO 19100 series of standards is 'capability to communicate, execute programs, or transfer data among various functional units in a manner that requires the user to have little or no knowledge of the unique characteristics of those units' (ISO/IEC 2382-1:1993). As correctly reflected in Figure 5: the metadata in the presented approach is accessible and searchable from different GIS packages (i.e. syntactic interoperability), but the other packages do not 'understand' it, so the metadata cannot be updated (i.e. no semantic interoperability).	The reviewer raises a good point here (similar to that of Reviewer 1 above). The approach described in the paper is interoperable in the sense that the automated elements of the metadata are updated when the data is edited. However, further work is required to ensure that the 'manual' elements of the metadata creation and edit process are made interoperable. Text has been updated to reflect this.
The authors describe their work as applying to an academic context, but this academic context is not described; there is no information to explain why the academic context is different to others? For example, in section 3.2 they explain that a significant number of metadata elements may be automated in an academic context, but they do not justify why this is different from other contexts. I could only find information in the last paragraph of the paper about the difference of an academic context, i.e. adding work package information. My suggestion would be to remove the constraint of an academic context from the title, abstract and paper; the work that is described could equally well be applied and used in a non-academic context.	Agreed. The text is now clarified to state that in an academic context datasets are usually created as part of a specific research project, from which the required information can be derived automatically through pre-configuration. This may not be possible in the more general case where the sources of datasets are not as constrained. Additionally, data curation is now important in an academic context due to requirements of research funders.
The keyword generation and dataset language detection described in section 5 are probably the most interesting automation features that the authors have implemented. These should be reflected in the objectives of the paper.	Agreed. The objectives have been updated to reflect this comment.

The title seems to suggest that there is a specific reason why FOSS4G was used, but there is no justification in the paper. For example, were the free PostGIS and QGIS installations a motivation or the fact that it is easy to write plug-ins, which may then be distributed freely with the software installation? This justification would also be interesting for the FOSS4G2013 target audience.	Agreed. FOSS4G was used as this makes the resulting approach freely and widely accessible to academics and students, without incurring licensing costs. The paper has been edited to explain this.
Finally, making use of triggers is an interesting tightly coupled approach, but triggers have their drawbacks and the authors should acknowledge this. For example, what will happen if a dataset of millions of points of interest is reprojected and each individual point of interest record is updated in the database? Will the last revision date in the metadata be updated for each point of interest? Moreover, moving the metadata updating functionality to the database hides the functionality from the user, which could result in unplanned side effects (such as a cascading effect). How do you plan to address these downsides of triggers in future?	We thank the reviewer for these very useful comments and have incorporated our responses to them in the 'future work' section of the paper.
3. Automating Metadata Creation 3.2 I suggest that you left-align the first column of Table 1 for better presentation.	Agreed. Text is now left-aligned.
4. Implementing Metadata Creation in FOSS GIS Figure 4: Replace 'Futures' with 'Future'	Agreed. Change made.
The text at the beginning of 4.3, before 4.3.1, should be moved into its own sub-section with an appropriate title.	The text provides a general introduction to the section, with the sub sections focussing on specific aspects of the system.
4.2.1 This subsection lists metadata details that are added. It is not clear when they are added? When the data is imported? Or when an update is made? The answer seems to be provided just before 4.3.1. This is too far away.	Agreed. Section 4.2.1 now forward-references section 4.3.1
4.3.1 It is not clear why the metadata language needs to be detected if the user has the opportunity to change the dataset title, abstract and lineage. Why don't you just add a dropdown for the language onto the dialog displayed in Figure 6? I agree that detecting the language of the dataset itself is a relevant challenge.	Agreed. Text clarified to indicate that the aim here is to minimise the manual metadata creation required by the user.
Testing Metadata Automation The footnote on p16 is important enough to be included in the main text.	Agreed. Text now in main body.
For each keyword in the third column of Table	Agreed. Text clarified to explain that this is

2, a figure is displayed in brackets. It is not clear what this figure is. Since the value for the fourth column is the same for all datasets, consider removing that column (the information is provided in the text), which will allow more space for the third column.	the number of matches for the keyword. Fourth column removed.
5.0.2 (p5) This subsection number is incorrect. Last paragraph: '…boundary extents and polygon for the dataset was updated…' should be '…boundary extents and polygon for the dataset were updated…'	Agreed. Section re-numbered and text changed.
ISO standards cited in the text are not listed in the references. They have to be added, e.g. ISO 19115:2003, Geographic Information -- Metadata. International Organization for Standardization, Geneva, Switzerland.	Agreed. Citation added.
2. References are cited incorrectly in some places, e.g. 'Beyond these basics, (Kalantari et al. 2010) have introduced…' should be 'Beyond these basics, Kalantari et al. (2010) have introduced…'	Agreed. However, this formatting is generated automatically by the template provided.
There are a number of incomplete references in the list, e.g. - Ellul et al. (2012): Who is the publisher of the book? Editors of the book? - Batcheller (2008): The volume and issue numbers are missing. - Deng & Di (2009): The volume number is missing. - Poore & Wolf (2010: No source is provided for the article	Agreed. Some issues caused due to bibtex referencing.
Language The article is well-written and easy to follow. One small issue: it is not clear why some words appear in quotes, e.g. 'catalog' (first paragraph, p5) and 'properties' (last paragraph, p5) are in quotes. I suggest to remove the quotes.	Agreed. Quote marks removed.

Reviewer 3

Feedback	**Comment**
This is an interesting effort to solve the "non-completion" problem in metadata through automation.	We thank the reviewer for his/her feedback.
I don't think you need all that information displayed in Table 2 to make your point.	Agreed. Both Tables 1 and 2 have been shortened and full versions made available online.
Read through your paper once more. There are a few typos.	Agreed. Text has now been corrected.