Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2016) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

13th International Conference on Current Research Information Systems, CRIS2016, 9-11 June 2016, Scotland, UK

Research information standards adoption: Development of a visual insight tool at the University of Cambridge

Mathias Riechert[a,b], Owen Roberson[c], Jürgen Wastl[c]

[a]*German Centre for Higher Education Research and Science Studies (DZHW), Schützenstraße 6a, Berlin, 10117, Germany*
[b]*Technical University of Berlin, Straße des 17. Juni 135, Berlin, 10623, Germany*
c University of Cambridge, The Old Schools, Trinity Lane, Cambridge, CB2 1TT, United Kingdom

Abstract

While Research Information continues to mature as an area of expertise, discussions regarding the implementation and adoption of standardisation initiatives, such as CASRAI and CERIF, have intensified. Possessing the capacity to use a standard does not obligate its adoption, so the extent to which standards are employed varies across use cases and institutions, in a way that is difficult to qualify and determine. We are presenting a light-weight visualisation framework for presenting, comparing and improving the adoption of research information standards for research institutions. The framework is implemented for providing insight into identifier adoption at the University of Cambridge. The tool is easy to deploy and implement, and the insights it generates are intended to express clearly the extent to which Research Information standards have been adopted. Furthermore, the framework can be used to make this adoption knowledge available as linked open data held at a local level, reducing the need for costly metastudies and helping the standardisation community to monitor and focus standardisation development.

Peer-review under responsibility of the Organizing Committee of CRIS2016.

Keywords: research information standardisation; research information system usage; standard adoption; visualisation; visualisation prototype; identifiers; linked open data

1. Introduction

In recent years, the conversation regarding the implementation of standardisation initiatives for Research Information (RI) has gathered pace and intensified. Although modern research information systems are compatible to research information standards, local decisions regarding the trade-off between usability and standardisation

Peer-review under responsibility of the Organizing Committee of CRIS2016.

creates an environment in which research information can be stored in formats that do not conform to the standards. This leads to three major challenges for the managers of research information systems and the standardisation community:

1) The promise of CERIF and the initiation of a UK CASRAI chapter provide two knowledge architectures to be considered when developing in this space, and companies involved the sphere of RI have invested attention and some development in the area of identity management, accommodating an array of external identifiers (examples of identity management for individuals include ScopusID, ArXivID, ResearcherID, and more recently ORCiD). With the exception of ORCiD, the other IDs are either explicitly subject-specific, or are implicitly so through the focus of coverage of the vendors that supply them. As disciplinary boundaries come under attack from new work practices while RIS admins are seeking ever-greater aggregation of data across the full range of the activities of their institutions, the subtle differences in the logic of how each identity works, and the areas it covers, will likely provide a productive site for future development. In order to identify such opportunities RI managers and administrators need to be able to get a view on the extent of penetration, the marginal benefit of penetration, and the extent of possible penetration of each identity. Our visualisation prototype enables RI managers to get an insight into the degree to which the identities are adopted across the institution.
2) Delivering insight from the large quantity of amassed RI data is hindered by deficiencies in data depth and breadth, which are being exposed as the complexity of queries grows. These deficiencies are caused by a range of factors including, but not limited to: poor user engagement (whether by real or perceived complexity in the systems, a weak understanding of the benefits, an ideological objection to such systems, or an irritation at duplication of requirements across different policies); uneven, and locally poor, coverage of disciplinary range amongst data providers; poor data management practices; and an unclear understanding amongst RI managers of the relative incidence of all these and other factors. We anticipate that the contribution of the visualisation architecture will provide benefit to two key groups:
    a. RI managers will be able to track the adoption of identifiers across the RIS over time, and see the impact of awareness-raising initiatives, such that they gain more understanding of the attributes of different areas of the user community and can better target advice and advocacy.
    b. Researchers will be exposed to more information about identifiers and the raised profile of these may help to increase awareness and encourage adoption of identifiers such as ORCiD in the RIS.
3) The third challenge such a project will seek to address is the present lack of information about the extent to which RI standards have been adopted by the research community. More clarity on this issue would enable the research information standardisation community to prioritise the interoperability agenda to an appropriate extent, identify technical and policy challenges effectively without recourse to time-consuming and difficult surveys, and provide an opportunity to RI managers and administrators to explore more fully the potentials of data migration through more widely adopted standards.

To address these challenges, we propose a light-weight-visualisation framework for presenting, comparing and improving the adoption of research information standards for research institutions. The framework is implemented for providing insight into identifier adoption at the University of Cambridge. Providing easy-implementable insight aims to increase awareness for the standardisation adoption of research information among researchers and managers. Furthermore, the framework can be used to make this adoption knowledge available as linked open data held at a local level, reducing the need for costly metastudies and helping the standardisation community to monitor and focus standardisation development. The remainder of the paper is structured as follows. Section 2 presents the visualisation framework and the used tools. In section 3 the main findings are given. Section 4 discusses limitations and section 5 gives a conclusion and discusses practical implications.

2. Related Work

Visualising research information has been part of all major commercial research information software systems since their beginning. Aside from research information software systems, Sci2 and Gephi have been used for the visualisation of research information[1]. Another application of visualisations in research information standardisation literature is to provide an overview over Research Infrastructures in Europe[2].

Extending the purpose of visualisations beyond providing insight, Dimou et al.[3] proposed improving the quality of linked open data on publishing by using visualisations and make them available for the public. Their approach primarily aims at enabling non-experts to explore the information available in linked open data. Further extending the purpose of using visualisations, Peña et al. presented a research information system providing visualisations and optionally publishing Linked Open Data as RDF[4] in parallel. All existing visualisations in the area of research information standardisation are built on top of research information like publications, projects, persons, etc., not about meta-information about the systems usage. In this study we focus on monitoring research information system data quality on a meta-level to improve standards usage.

The design of the visualisation prototype itself draws from information visualisation literature. In information visualisation research, reducing the complexity of visual information to allow for better information perception and to amplify cognition has been a central topic since the emergence of the field. A very prominent approach to reach this goal is the ‘visualization mantra’[5,6]. As existing tools approaches are focused on network analysis or require commercial software to render visualisations for web-usage, we implement the practices common to information visualisation literature on research information usage data in the following.

3. Approach

In order to provide a proof-of-concept that providing a visual insight tool is easy to implement with free tools only, we conducted a rapid prototyping phase of 4 weeks, in which the data loading and visualisation prototype was

developed. Figure 1 shows how data is loaded from the research information system and transformed for the visualisation.

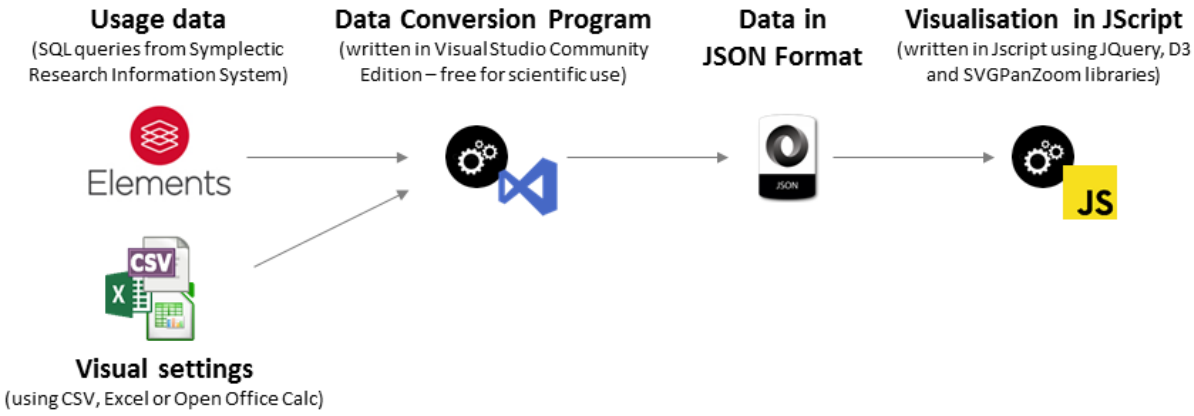

Fig. 1: Framework for data loading and staging

In order to provide a flexible and reusable system, we separated the data load process (usage data in the figure) from the visual settings. Usage data is loaded with SQL queries from the research information system used at the University of Cambridge. As discussed in the outlook, this can be interchanged with any usage data which can be accessed with SQL or by Linked Open Data. The visual settings provide the information of the coordinates and attributes of the elements representing parts of the research information system (see Figure 2 for an example of a visualisation showing the ORCID identifier as a node). The coordinates and initial visual settings are extracted from a previous version of the research information map of the University of Cambridge[7]. The transformed usage data in JSON is then displayed in the visualisation framework, shown in Figure 2.

The proposed framework is separated into a visualisation layer and a data layer to allow both components to be modified, exchanged and maintained independently.

In the visualisation layer, an SVG representation of the research information system is rendered based on node coordinates and information provided in Excel (or openOffice Calc). In the example of the RIS of the University of Cambridge, the red line represents all data included in the RIS Symplectic Elements. In the 4 week prototyping phase, we focused on the Identifier line (green, connecting all ID's used at the University of Cambridge). It is important to highlight that the approach is not restricted to identifiers but rather can be extended on complete data formats (like CERIF) or on RIS contents with another visual layer.

By using SVG the zooming interaction to the Identifier level allows the user to focus on the line while blending out surrounding elements without a reduction of graphical quality. By clicking on a node in the graph (representing the ORCID Identifier in the example) the user zooms to the detail level, on which the usage percentage is presented by circle segments and other elements are filtered out. Multiple percentages can be displayed allowing comparisons through cross sectional or time series data. In this case we are looking at the changes in adoption over time.

Additional information is provided and links to other differentiated graphs are provided on the right. Furthermore, links to websites or internal information portals are added here.

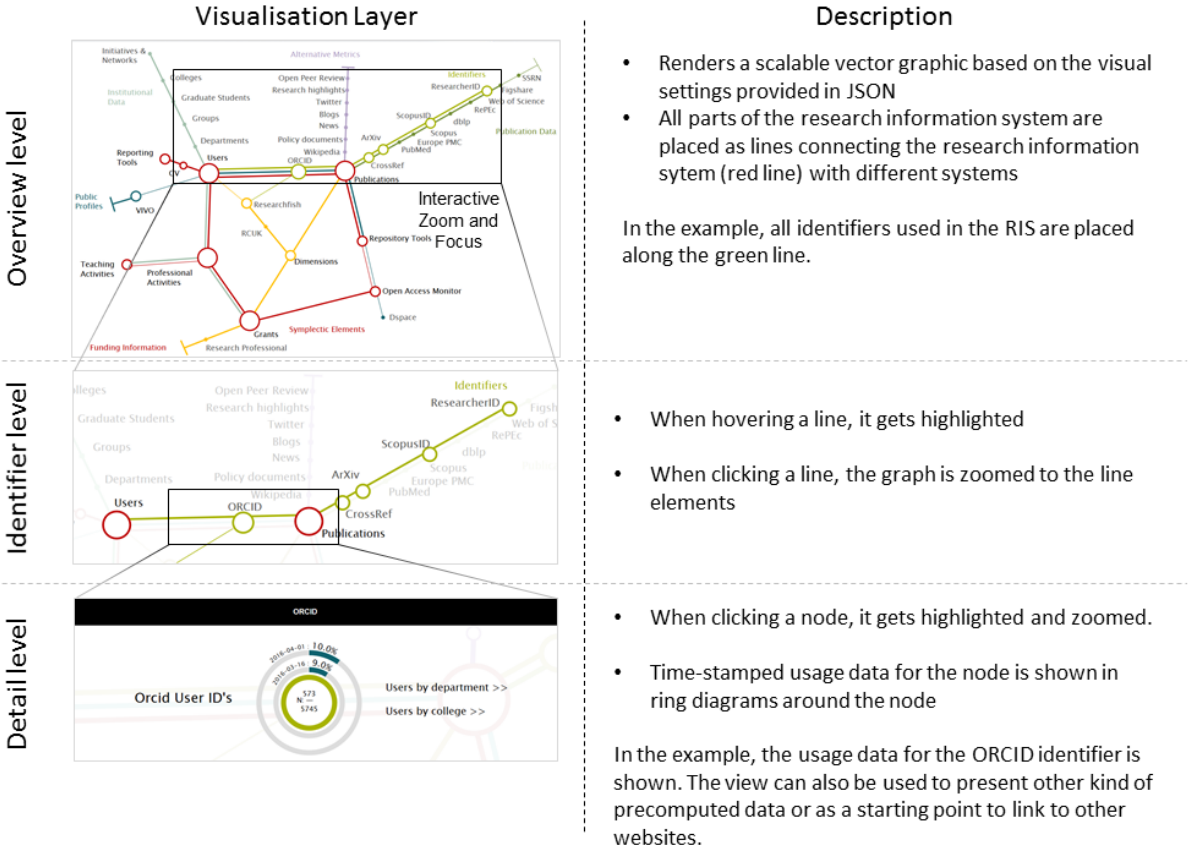

Fig. 2: Visualisation concept for Research Information Standardisation Adoption

The underlying meta-information about the adoption percentages is managed in the data layer. The data is extracted by a simple tool that uses customizable SQL queries to access and aggregate data from the underlying RIS (Symplectic Elements in the example). The tool then transforms the data to a time-stamped light-weight JSON file to allow data to be collected as a time series. We are considering an open access licence for the tool (it could as well be implemented easily as an export by software vendors). By using the identifiers specified in the SVG and adding classes for groups of elements (like all elements along the green indicator line in our example) those groups can be made interactively focusable. In the JSON, each node contains one or more percentages which hold the precomputed usage information. This allows for flexible annotation with any kind of information or link that is to be displayed with the node.

The two layers are integrated in the visualisation frontend. It is programmed as a light-weight web application (HTML5 + Javascript), which requires no installation and can be embedded easily in other online profiles like VIVO or institutional websites. Interaction, zoom and context information functionality is added to the SVG using the IDs

and classes in the JSON. This allows for different research institutions to use the same graphical layer but to repurpose the supporting metadata fields for their institutional and business context. We strongly encourage subsequent users to consider the same set of node identifiers and classes across different research institutions as this would result in the possibility of aggregating JSON files across research institutions and enable global standardisation monitoring.

In order to demonstrate the feasibility of the prototype, we conducted a four-week data integration phase. In this phase, we analysed which Identifiers could be computed or integrated from sources already connected on another part of the RI network. As an example we analysed the ORCiD, ArXiv, ScopusID and ResearcherID profiles, the latter three of which had not previously been promoted or tracked in the RIS at the University of Cambridge. By creating timestamped JSON data, we could in parallel show the changes in the percentages over the period of the four weeks.

4. Findings

The rapid prototyping showed that providing a visual portal for monitoring adoption of identifier adoption can be implemented online in 4 weeks. The architecture, with a separation of the visual and data layer, allows for cross-institutional and cross-context reuse such that these techniques can be applied in the general case beyond identifiers to standards adoption. By using only freely available tools, this is achievable for all research institutions without the need to buy cost-intensive licences. The only prerequisites are an SQL or SQL-like source with the information about users and their identifiers, and somewhere to host the tool. As such, the barrier to entry is very low, and research institutions without a commercial RIS are enabled to visualise and document their standards adoption as long as the data can be aggregated. The visualisation portal can also be extended to link to other front-ends and portals and can as well be embedded in them.

By conducting a parallel data integration phase we could also show that the approach is feasible for presenting changes in the standards adoption. Scaled-up to the whole research system this would enable analysis and visual presentation of standard usage on a scale limited only by its adoption by the community. We also attach the used visual settings for coherent usage on scale-up.

In comparison with existing visualisation approaches, our approach is the first to visualise usage data of research information adoption. It is of course possible to use commercial or embedded visualisations for that pupose. By using only freely-available tools we developed an easy, aesthetically pleasing and user-centred prototype, which can be extended with very limited knowledge of visualisations in D3 or JScript to show whatever graphs are required or functions are wanted. In contrast to existing commercial approaches, we are not creating a rendered picture for web-access, but have all visualisations component computed based on visual settings and pre-calculated data. This approach is therefore flexible enough to accept different input sources and generate different output formats without framework restraints.

5. Limitations

Due to the exploratory nature of the development process, no quantitative evaluation of the prototypes perceived usability, aesthetics, usefulness and ease of use have been conducted at the present stage of development. While developing the prototype, the head and members of the Research Information Office of the University of Cambridge were included in qualitative feedback meetings to align the data flow and interface along research managers' requirements. In this qualitative phase it is too early to generalise from the department member's view of the prototype being useful to the whole university. This will be addressed in a subsequent quantitative phase, where the visualisation is exposed and evaluated by research managers, researchers and the research information standardisation community.

At present, our prototype only shows data for research information identifier usage. Although the prototype is designed to display any kind of percentage along the nodes, increased value could be realised by having more differentiated data appended to the graph. For example, it would be interesting to differentiate percentages by department or organisational unit in order to address departments with low usage rates directly. This is now implemented by having a link to a differentiated view on the detail level. Given more time and effort, implementing differentiated visualisation approaches based on disaggregated data would be beneficial for showing this kind of information directly.

In order to apply this approach to other research institutions, limited programming knowledge in the visualisation domain is needed. Once this comparably small learning hurdle is overcome, a department implementing such visualisations is rewarded with boundless flexibility concerning visualisation contents, interactivity and functions. This enables them to provide user-centred visualisations aligned to their organisations' reporting and information requirements.

6. Conclusion and Outlook

The proposed framework provides an opportunity to contextualise the generally normative management-level conversations about data standards adoption with evidence that is trivial to gather and easy to interpret. It complies with the letter of the standards and the spirit of the open data concept on research information usage in RIS. We believe that this represents an important step on the way to a better understanding of how research information is documented and perceived.

A novelty of our approach is the focus on freely-available, simple and widespread web-technology with flexible expansion and re-usage possibilities and minimal implementation and maintenance effort. It is our intention that the ready reproducibility of the pilot will encourage other research institutions to join the conversation in this space and help manage its future direction.

In future research and development, three main routes will be taken:

Firstly, we will enable the transformation program directly to provide linked open data of the usage percentages. Although this part is planned to be optional, we believe researchers in the field of research information standardisation and research managers would benefit if this usage knowledge was readily and freely available on the web. By a widespread adoption of this approach the need for time- and money-intensive studies on the usage of research information standards could be reduced by large parts.

Secondly, we plan to evaluate quantitatively the prototype's perceived usefulness, its ease of use, its acceptability and its aesthetics with respect to target users of the visualisation. By widening our current qualitative evaluation of the prototype to larger parts of the research administration and researchers, we hope to address the awareness goal for indicator usage, and to develop a more generalizable understanding of how far this visualisation supports the goals outlined in the introduction. In a later step, comparative studies across different research institutions and national research contexts could help to differentiate that knowledge to different organisational and national cultures.

Thirdly, we plan to implement on an explorative basis extended kinds of visualisations based on this framework. Those could be implemented as switchable modules to provide a consistent set of different perspectives on the differentiated usage data of research information.

7. Acknowledgements

The authors wish to express their thanks to the German Academic Scholarship Foundation for funding the research stay.

8. References

1. Dijk E. Visualization of NARCIS: a network of experts and knowledge organization in the Netherlands. 2012;
2. Houssos N, Karaiskos D. The CERIF-based EuroRIs-Net+ Observatory. 2013 Nov 15;
3. Dimou A, De Vocht L, Van Grootel G, Van Campe L, Latour J, Mannens E, et al. Visualizing the information of a Linked Open Data enabled Research Information System. In euroCRIS; 2014.
4. Peña O, Aguilera U, Almeida A, López-de-Ipiña D. Labman: A Research Information System to Foster Insight Discovery Through Visualizations. In: Garoufallou E, Hartley RJ, Gaitanou P, editors. Metadata and Semantics Research. Cham: Springer International Publishing; 2015. p. 286–97.
5. Shneiderman B. The eyes have it: A task by data type taxonomy for information visualizations. In: Visual Languages, 1996 Proceedings, IEEE Symposium on. 1996. p. 336–343.
6. Shneiderman B, Plaisant C. Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies. In: Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization. 2006. p. 1–7.
7. Wastl J. Mind the (Research Information) Map [Internet]. 2015 [cited 2016 May 20]. Available from: https://figshare.com/articles/Mind_the_Research_Information_Map/1507607