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

# Developing and Standardising Definitions for Research Information: Framework and Methods of Successful Process Documentation

Mathias Riechert[a,b] , Sophie Biesenbender[a], Christoph Quix[c]

[a]*German Centre for Higher Education Research and Science Studies (DZHW), Schützenstraße 6a, 10117 Berlin, Germany*
[b]*Technical University of Berlin, Straße des 17. Juni 135, 10623 Berlin, Germany*
[c]*Fraunhofer Institute for Applied Information Technology FIT, 53754 St. Augustin, Germany*

**Abstract**

We present a framework for a transparency-oriented process of developing semantic definitions for research information. The framework is applied to the project "Research Core Dataset" (2013-2015) that provided the context for developing a set of core definitions for research information for the German science system.


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

*Keywords:* research information standardisation; argumentation visualisation, incremental definition development framework

## 1. Introduction

Agreeing on standardised definitions of research information (RI) – for example how academic staff, publications, patents or funding are to be counted – is fundamentally important for a commonly agreed and fair interpretation of the reports created on the basis of this information. As those reports are increasingly used to inform performance assessment and even funding allocation, reaching explicit agreement on these RI definitions has become more and more important in output-oriented science systems.

Recent studies in standardisation literature on research information (RI) show that the process of reaching a consensus about a semantical definition is vitally important for the use of standards like CERIF or CASRAI. Although many projects have shown that RI standards like CERIF or CASRAI are already vastly used on institutional level, agreeing on the underlying definitions for these standards still remains a central challenge for the interoperability of different research information systems (RIS)[1]. Different ways, in which a standard can be mapped might decrease the potential power of the standard itself[2] and lead to the rejection of the standard altogether. If we

1877-0509 
Peer-review under responsibility of the Organizing Committee of CRIS2016.

consider the process of semantically defining research information as a "wicked problem"[3,4], central solution strategies from the "wicked problem" literature can be applied to the process of RI standardisation.

Acceptance of the solution among the stakeholders is of central importance for addressing wicked problems (such as for example agreeing on definitions in RI standardisation). For this reason, Kunz and Rittel[5] developed Issue Based Information Systems (IBIS) to enable visual representation of argumentation processes that lead to an agreed solution to the discussion. By providing a visual representation, IBIS aims to ensure a transparent working procedure. It explicitly documents alternatives and arguments in order to focus the discussion and to more easily reach agreement and acceptance.

We embed the use of the recent IBIS tool *Compendium* in a RI definition process, in which the number of participants is gradually increased. To avoid redundant discussions during this process, documentation of the results, but also of the alternatives and arguments is a key point of our approach. As the complexity of information is quickly growing during the definition process, appropriate visualisation techniques are required[6]. We discuss the framework in the context of the project "Research Core Dataset" (RCD) for the German science system (2013-2015). The project provided the context for developing definitions for a set of core RI definitions[7,8]. The process was initially started by a working group of the German Council of Science and Humanities. Detailed discussions on RI definitions followed within several working groups dedicated to specific areas of RI (e.g. publications, subject classification). The developed definitions (along with the documented discussions) were then reviewed by an advisory board – followed by a public request for comments. More than 100 stakeholders and organisations from the German science system made use of this opportunity and submitted feedback on different definitions.

We reflect upon our experience and insights gained through accompanying research of these processes. The goal is to provide best practices for similar projects and initiatives in the RI standardisation community and to improve the way in which definitions are developed. A related paper discusses the modelling process and the mapping of the data model to the CERIF standard[9].

## 2. Related Work

*Related work research information standardisation literature:*

It has been shown that the process of reaching consensus about a semantical definition is vitally important for the use of research information standards like CERIF or CASRAI. Yet, interoperability remains a central challenge [1,10]. This becomes evident in cases where the direct mapping to standard CERIF semantics is not possible due to diverging definitions of the concepts. Mapping CERIF is often a trade-off between the implementation of individual definitions and the interoperability between different systems[2]. The varying ways a standard has been mapped might compromise the potential power of the standard[2] and diminish the usefulness of the standard altogether. In this context, Sidselrud and Lingjærde[11] highlight the importance of broadly accepted definitions and standards[7,8].

Riechert and Dees[3] discuss and interpret the process of semantically defining research information as a "wicked problem". The complexity of agreeing on a definition is also reflected in the use of CSAV for RI standardisation.

*Related work in Computer-Supported Argumentation Visualisation literature:*

Computer-Supported Argumentation Visualisation (AV) roots in wicked problem research. This way, it extends Kunz und Rittel's[5] Issue Based Information System (IBIS) notation. The most recent tool for modelling in IBIS is Compendium[12]. Numerous studies have since been published on the influence of argumentation visualisation on the working atmosphere in discussions (see e.g. Schneider et al.[13], Scheuer et al.[14], and Suthers[15] for an overview). The most important effects found in quantitative user studies are the following.

- AV increases the discussion about the links between different aspects and issues of debate[16].
- AV increases the degree to which past topics of discussion are taken up in the on-going discussion [16,17].
- AV reduces the time for a hypothesis to be formulated in group discussions[15].
- AV increases the number of discussed alternatives[15].
- AV reduces the time for an alternative to be agreed upon[15,17,18].
- AV increases the flexibility to change participants' positions in case an argument is convincing[18].
- AV reduces the number of times a certain discussion topic is discussed redundantly[17].
- AV helps keeping the focus on the discussion topic[19].

Although studies on the use of IBIS maps for communicating discussion contents found that argumentation maps have advantages over textual representation[20,21] our experience with detailed argumentation maps is mixed. Incorporating 300 RI definitions entails high graph complexity, which in turn causes serious issues for graph presentation, understandability and usability. These circumstances increase access barriers. We therefore developed new forms of representation to allow for more structured illustration of the complex development process.

*Related work in incremental software development literature:*

Instead of using a waterfall-like model to develop definitions for research information, we draw on the spiral model as described by Boehm[22]. Incremental development models allow for immediate and direct user feedback on the present stage of development in each iteration. For higher scalability, the spiral model can be further formalised by using the SCRUM model[23], which focuses on development projects in contexts with constantly changing requirements. As SCRUM requires formalised rules and professionalised staff, we referred to the spiral model as the basis for the development of definitions in the context of the RCD project.

*Added Value of our approach:*

To the best of our knowledge the application of CSAV (in the form of IBIS modelling) has not been discussed in RI standardisation literature yet. By embedding CSAV in a framework for the integrated development of RI definitions we were able to reach consensus on 300 definitions and sub-definitions. The process involved the main stakeholders of the German science system. This was done by extending existing approaches of representation, developing new forms of visualisation and linking them to a conceptual data model[9]. This way, we enabled both directly involved and external actors to participate in the deliberation and the decision-making process. Applying this approach to standardisation projects in other countries or international initiatives might create more systematic knowledge on RI standardisation and the agreement on definitions (including alternatives and arguments). Finally, we hope to support interoperability by proposing a framework which also reflects on underlying argument processes for different deliberation and decision-making contexts.

## 3. Approach

In the following we propose a strategy for solving wicked problems in processes of defining research information, i.e. Computer-Supported Argumentation Visualisation (CSAV). Based on this approach, we document the closed discussion processes within expert groups including the discussed definitions, alternatives and arguments. The expert discussions were subsequently made available to an increasing number of stakeholders. Additionally, the stakeholders and the public were given the possibility to retrace the discussion process and to comment on the results (i.e., the definitions) based on the CSAV output.

### *3.1. Using CSAV for modelling definition processes*

In CSAV literature, Issue-Based-Information-System (IBIS) notation is the most commonly used way to model and document argumentation processes. Figure 1 illustrates the basic elements in IBIS.

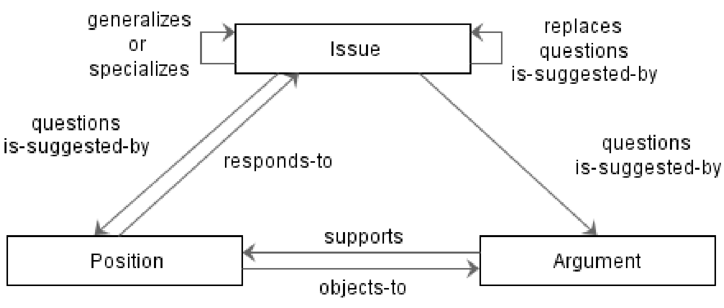


Fig. 1: IBIS elements[24]

In our case we used *issue* elements for the question “What is a suitable definition for …?”. The *position* element represents the respective answer (i.e. definition) and its alternatives. *Arguments* are being connected to the position to document arguments raised in the definition process. As a tool for modelling the definitions in IBIS we used Compendium[12], which allows for fast and intuitive graph editing, tagging and export of argumentative models. Figure 2 shows how Compendium is used to document the definition process for research information. The definitions are saved in the position (or answer) node itself and all nodes are connected as defined in IBIS language.

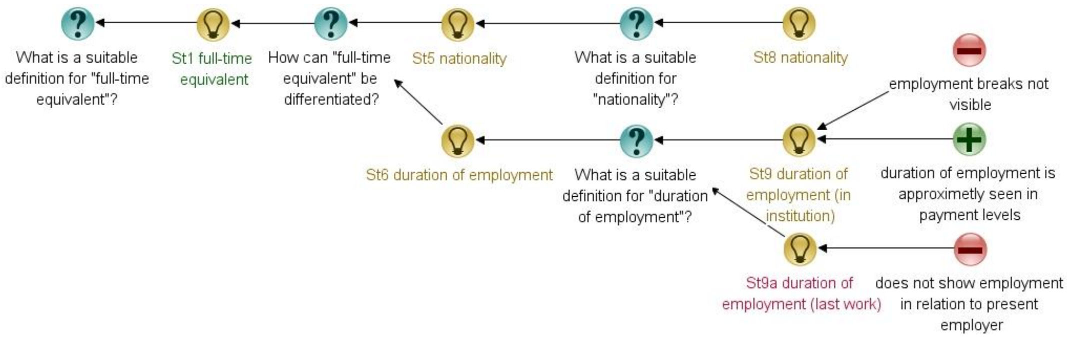


Fig. 2: Example of using Compendium to model definition processes in IBIS[6]

Since definition networks get very huge in real world applications, we further simplified IBIS for modelling definitions. Figure 2 presents an excerpt of the discussion in IBIS notation. Although Compendium allows for free placement of questions (issues) and answers (positions), we introduced a few further rules to increase readability, reduce graph complexity and allow for automatic re-usage of the argumentation model. The semantics and formal language will be discussed in detail in a separate article which is currently in preparation.

In line with CSAV literature, modelling the definition process on-screen while discussing definitions with the expert workgroup (contrary to ex-post modelling) has a number of benefits, which are discussed under in the sections on *findings* and *practical implications*.

*3.2. Embedding CSAV into a framework for Stakeholder-Integrating Development of Definitions (SIDD)*

SIDD (Figure 3) was implemented to provide insights on the discussion process and in order to give participants the possibility to take part in the early development process in a systematic way.

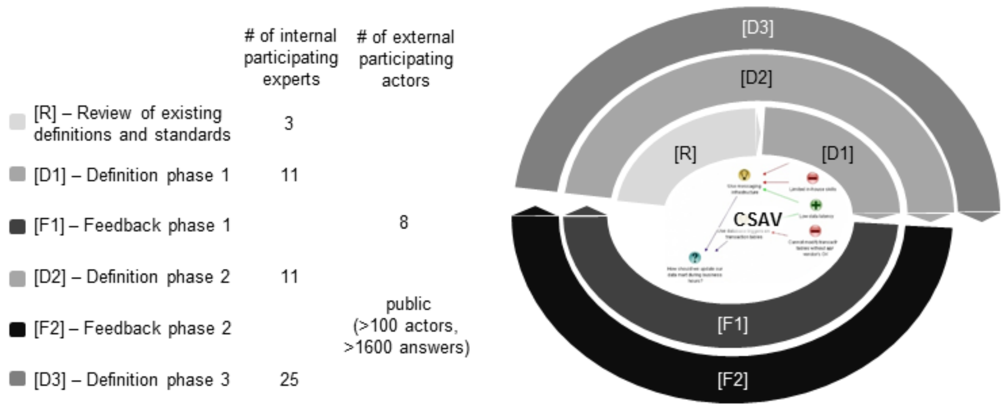


Figure 3: CSAV-Centred Incremental Definition Development Framework

The definition process was designed as an incremental cycle model, in which phases of definition development by closed expert groups alternate with phases of presenting the state of definition to external experts. With each iteration, the number of participating internal experts (being responsible for defining and editing based on external feedback) and external actors (being responsible for reviewing the draft, pointing to problems and suggesting alternative definitions and arguments) increased.

Using Computed Supported Argumentation Visualisation (CSAV) as described in the previous section is the core of the framework. All **definition** phases (R, D1, D2, D3 – depicted in grey colour and located on the upper side of the cycles) were conducted with the help of on-screen modelling of the definitions, alternatives and related arguments. Before each discussion session, the model was prepared by the discussion moderator. In line with CSAV literature, her or his role goes far beyond moderating, which is why the term 'discussion facilitator' is more appropriate. In addition to preparing the discussion model the facilitator also prepared the questions that followed up from existing definitions and standards or feedback phases. All **feedback** phases (F1, F2) used forms of visualisation and presentation from the argumentation model to provide consistent insight into all definitions, alternatives and arguments discussed. This allowed for a structured presentation of the discussion and its results. For this reason, the model was extended and used as a platform for gathering structured feedback from external experts and stakeholders. The arguments documented reflect expert opinions and the state of discussion. In order to allow for open and topic-oriented discussion, the arguments were documented anonymously. The weight of the arguments and the eventual definitions were subject to change over different definition and feedback phases.

### *3.3. Forms of presentation and visualisation used for documentation and obtaining structured feedback*

We consider documenting past and on-going argumentation processes as central for integrating external stakeholders' knowledge, alternative definitions and arguments into the definition model and for increasing acceptance for the standardised definitions among all stakeholders. In order to transform the argumentation model and the respective definitions, we implemented a process and developed visualisations as shown in Figure 4.

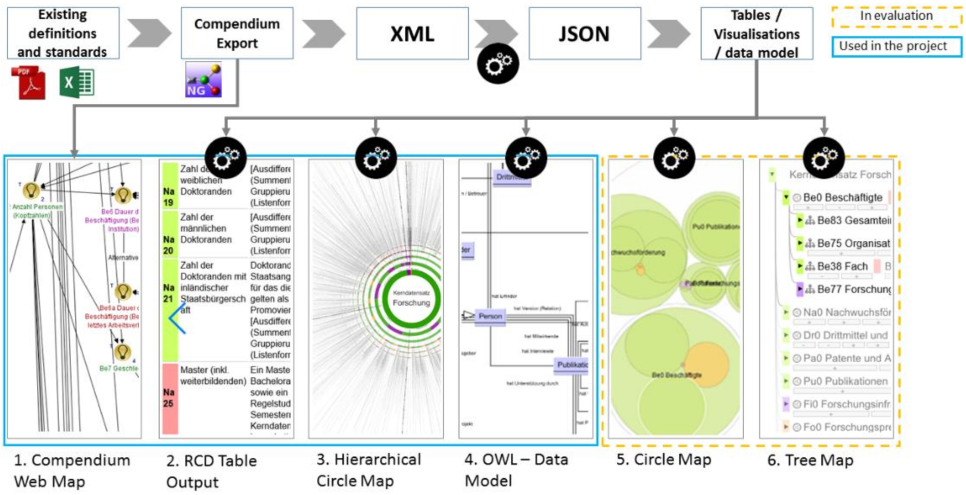


Figure 4: Process Architecture and Developed Visualisations

Compendium features a Web Map (1.) export, showing the argumentation model as a network graph with definitions available on click. As our experts in the project were overwhelmed by the complexity and number of nodes and connections (at the end of the project we had >300 definition nodes, 250 alternatives nodes, 600 argument nodes in that graph) we developed complexity-reducing interactive representations based on the modelled definition network. First, we exported the model data to XML. A custom written program (symbolised by the black circle with gearwheels) transformed this Compendium XML to the hierarchical JSON format. This was then used as the basis for five different forms of representation described below:

- 2. RCD Table: this form of representation shows all definitions, alternatives and arguments in an interactive, sortable and printable HTML list. To increase clarity, we added an interactive index table that allows individually expanding and reducing definitions in lower hierarchy levels of the definition tree. This table was also used for gathering public feedback (see previous section). During this phase, external stakeholders were invited to add comments, alternative definitions or arguments for each definition. The feedback tool allowed for a very precise and structured feedback process. In total, we registered more than 1800 comments on the definitions.

- 3. Hierarchical Circle Map: although it is possible to show the whole hierarchy (circle segments) and all definitions (rays) in one scalable vector graphic, this visualisation was primarily used for presentations and logos.
- 4. OWL Data Model: the definitions developed in the previous steps were used for the development of a conceptual data model for implementation in RIS[9]. The structured documentation of definitions and arguments was helpful for the development of the data model because it clarified the interpretations of the definitions. The data model itself is formalised as OWL ontology and documented in a web-based system which links model elements with their corresponding definitions and arguments. The data model was updated continuously according to the changes in the argumentation model in subsequent iteration phases. Furthermore, we linked the data model to CERIF by implementing the conceptual data model in an XML-based data exchange format.
- 5. Circle Map / 6. Tree Map: these two forms of visualisation represent exactly the same information as the previous forms of visualisation, though in an aggregated form to reduce perceived complexity. Both forms of visualisation allow for user-driven navigation through the definitions, their alternatives and arguments.

Visualisations 1-4 were already used during project period. Visualisations 1, 2, 3, 5 and 6 as well as additional process and protocol representations are currently being compared and evaluated in an on-going evaluation study. At the conference, we would like to show and discuss different forms of representation of RI definitions.

Summing up, we used CSAV as a basis for modelling RI definitions, alternatives and arguments. We embedded the technique in an incremental development framework with alternating development and feedback phases. Following that, we developed and used different forms of representation to make the structured definition content available for large audiences. The whole framework is designed to be scalable to larger (or international) projects for developing definitions of different kind. For this purpose, all the steps taken in the RCD project were documented and made accessible in a structured and incremental way.

## 4. Findings

By using this framework, three main achievements were reached in the project:

- In line with the literature on Computer-Supported Argumentation Visualisation (CSAV) like IBIS and its application in group meetings, our case reveals that the discussions and processes to agree on definitions (where experts discussed the definitions) became more structured and less status-biased. Overall, reaching consensus among the experts was faster. Furthermore, participants were able to internalise the state of the definition more effectively compared to a situation without on-screen modelling. The experts in our definition groups were initially a bit reluctant (not being used to this way of discussion support), but soon referred to the visualisation during the debate. Additionally, modelling the discussion on-screen observably improved the discussion performance in later meetings because the state of discussion and arguments for the different definitions were readily and quickly available for all experts. A further important improvement is the long-term availability of the discussion content. We expect all standards to require revisions from time to time. The composition of the group of participating experts is expected to change. Being able to draw on not only definitions but also alternatives and arguments in a structured form allows new experts to understand and reconstruct past decisions more efficiently compared to a discussion that starts from scratch. Finally, exporting and re-using the discussed contents by means of data models, definition tables, visualisations or any form of representation in an automated way reduces the documentation costs that emerge from changes in the definitions.
- The framework we propose and used in our project applies agile software-engineering principles to improve the way in which RI definitions are developed. Implementing these principles arguably required adjustment of the project processes. Firstly, making the rationales for and against different definitions more transparent and public in order to broaden legitimacy of the decisions taken, stands (at least at first sight) in contrast to the expert-centred approach, which was originally advocated by the German Council of Science and Humanities for use in the RCD project. Secondly, this way all major stakeholders of the science system and a huge number of research institutions could be integrated in the project and contributed to structured decision-making and deliberation

processes. Of course, such an approach comes at the cost of definitions that are to some extent watered down (if compared to the original ones agreed upon in the expert group). This stands in contrast to the gain in stakeholder acceptance of the definitions and the standardisation initiative in general. We expect this trade-off between broad standardisation and individual applicability of the standard to be true for cases beyond the national context as well. A lack of applicability might be mitigated by developing definitions for separate explicit business cases in order to limit the number of stakeholders to integrate into the decision-making process. The public feedback phase showed that a number of issues and definitions have not been addressed by the project so far. The comments also illustrated possible extensions to the RCD in future standardisation endeavours. Furthermore, the public feedback phase revealed that nearly fifty percent of the feedback was of positive nature, which was rather surprising because we expected participants to communicate criticism in the first place.

- Different forms of representation of the results (including definitions, discussed alternatives and arguments) provided insights into the underlying process of deliberation and discussion. Feedback on the project suggests that perceived complexity of the process is reduced by using interactive forms of representation. These tools allow external stakeholders to freely navigate through the definitions. In addition, the interactive representations applied in the project support linking and searching the content. In our experience, these representations are of high value to supplement traditional forms of documentation like protocols or plain lists of definitions.

## 5. Limitations

Using our approach in real large-scale standardisation projects comes at a cost: the RCD project involved three full-time researchers assisted by research assistants over a period of almost two years to organise communication, documentation, preparation and post-processing of the definitions as well as to develop and integrate the process workflow and transformation and representation software[7,8]. Although this framework is scalable to larger contexts (e.g. through parallel subgroups that work on the same argumentation model database), documentation efforts will increase with project size. Furthermore, a process like this might run much more efficiently with professional permanent staff (especially for project management, facilitation, modelling as well as model maintenance).

Our conclusions are based on a standardisation endeavour for the German science system, where the regulation of higher education institutions (HEI) falls under state responsibility. In addition, non-university institutes (Leibniz association, Helmholtz association, Max-Planck society and Fraunhofer society) play an important role within the science system. Research in these institutes is often interdisciplinary and organised along organisational units rather than academic disciplines, which makes it difficult to coherently classify research processes in HEIs and non-university research institutions. Therefore, RI standardisation in Germany requires agreement between a multitude of stakeholders and (state and federal) regulators. This process relies on consensus by deliberation rather than top-down decision making. For this reason, the present case constitutes a particularly complex case. Yet, we believe discussing the application of the used framework in different standardisation initiatives like CERIF or CASRAI may improve efficiency, scalability and reusability of the definition processes in these contexts as well.

## 6. Conclusion and Outlook

The framework presented in this paper was applied in a standardisation project of national scope. In this context, it helped a high number of diverse stakeholders to agree on a set of different definitions. Modelling the definitions, their alternatives and arguments in an argumentation model allows broad stakeholder groups to comment on and modify current decisions and definitions. The structured documentation of the definitions and their evolvement, their alternatives and arguments also facilitates automatic transformation into various forms of documentation to address different target groups with diverse levels of involvement, interest in the contents and backgrounds.

Future analyses of the RCD standardisation project are centred on the degree to which different visual representations are apt to enhance perceived transparency of the project as a whole and the decision-making processes in the RCD project. The evaluation takes currently place based on a survey with researchers and administrative staff in research-performing institutions.

Future analyses might focus on the applicability of the proposed framework in large-scale deliberation or standardisation settings, e.g. by documenting decision making in other environments (such as CASRAI).

The structured development of definitions also allows for embedding the discussion results, the underlying process and the developed representation in a larger “Development Transparency Framework”[6]. This framework focuses explicitly on increasing perceived transparency and will be subject to further studies.

Finally, the structured recording of statements allows for systematic content analysis. Also, definition matching is supported based on the structures to increase the interoperability of definitions in different contexts.

We like to initiate a discussion in the CRIS community about how agreement about definitions is being reached and on whether on-going efforts might benefit from an approach as proposed in the present framework. From our view, it enhances the structuring of the development process and provides the possibility to use knowledge about the arguments for or against different definitions. In the long run, this might support building a knowledgebase on alternatives and arguments of RI definitions. This information may also be used to increase RI interoperability.

## 7. References

1. Jörg B. Standardising research contexts towards system interoperability–and more. SK-CRIS Event. 2014;
2. Clements A, Brander S, McCutcheon V, Brown J, Heenan D, Vestdam T. CERIF in Action: Synthesise, standardise and productionise CERIF for Higher Education Institutions. 2012;
3. Riechert M, Dees W. Research Information Standardization as a Wicked Problem: Possible Consequences for the Standardization Process. Case Study of the Specification Project of the German Research Core Dataset. Procedia Comput Sci 33. 2014;272–277.
4. Riechert M, Biesenbender S, Dees W, Sirtes D. Developing Definitions of Research Information Metadata as a Wicked Problem? Characterisation and Solution by Argumentation Visualisation. Program [Internet]. 2016 [cited 2016 Feb 25];50(3). Available from: http://www.emeraldinsight.com/doi/abs/10.1108/PROG-01-2015-0009
5. Kunz W, Rittel HW. Issues as elements of information systems. Vol. 131. Institute of Urban and Regional Development, University of California; 1970.
6. Riechert M, Quix C, Zarnekow R. Fostering Transparency in Policy Development Processes-A Development Transparency Framework. 2015;
7. Biesenbender S, Hornbostel S. The Research Core Dataset for the German science system: challenges, processes and principles of a contested standardization project. Scientometrics. 2016;(106):837–47.
8. Biesenbender S, Hornbostel S. The Research Core Dataset for the German science system: Developing standards for an integrated management of research information. Scientometrics. 2016;1–12.
9. Quix C, Riechert M. Modelling National Research Information Contexts based on CERIF. In: 13th International Conference on Current Research Information Systems (CRIS). Scotland, UK; 2016.
10. Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, Read E, et al. Data sharing by scientists: practices and perceptions. PloS One. 2011;6(6).
11. Sidselrud A, Lingjærde GC. The practical implementation of the CRIS system CRIStin and the goals/challenges of bringing 150 institutions into production within a year. In: Proceedings of the 11th International Conference on Current Research Information Systems (June 6-9, 2012, Prague, Czech Republic). 2012. p. 305–312.
12. Selvin A, Buckingham Shum S, Seirhuis M, Conklin J, Zimmerman B, Palus C, et al. Compendium: Making meetings into knowledge events. 2001;
13. Schneider J, Groza T, Passant A. A review of argumentation for the Social Semantic Web. Semantic Web. 2013;4(2):159–218.
14. Scheuer O, Loll F, Pinkwart N, McLaren BM. Computer-supported argumentation: A review of the state of the art. Int J Comput-Support Collab Learn. 2010;5(1):43–102.
15. Suthers DD. Empirical studies of the value of conceptually explicit notations in collaborative learning. In: Knowledge cartography. Springer; 2008. p. 1–23.
16. Suthers DD, Hundhausen CD. An experimental study of the effects of representational guidance on collaborative learning processes. J Learn Sci. 2003;12(2):183–218.
17. Lund K, Molinari G, Séjourné A, Baker M. How do argumentation diagrams compare when student pairs use them as a means for debate or as a tool for representing debate? Int J Comput-Support Collab Learn. 2007;2(2–3):273–295.
18. Nussbaum EM, Winsor DL, Aqui YM, Poliquin AM. Putting the pieces together: Online argumentation vee diagrams enhance thinking during discussions. Int J Comput-Support Collab Learn. 2007;2(4):479–500.
19. Munneke L, Andriessen J, Kanselaar G, Kirschner P. Supporting interactive argumentation: Influence of representational tools on discussing a wicked problem. Comput Hum Behav. 2007 Mai;23(3):1072–88.
20. Loukis E, Wimmer M. A multi-method evaluation of different models of structured electronic consultation on government policies. Inf Syst Manag. 2012;29(4):284–294.
21. Renton A. Seeing the point of politics: exploring the use of CSAV techniques as aids to understanding the content of political debates in the Scottish Parliament. Artif Intell Law. 2006;14(4):277–304.
22. Boehm BW. A spiral model of software development and enhancement. Computer. 1988;21(5):61–72.
23. Schwaber K. Scrum development process. In: OOPSLA Business Object Design and Implementation Workshop. 1995. p. 10–19.
24. Conklin J, Begeman ML. gIBIS: a hypertext tool for exploratory policy discussion. ACM Trans Inf Syst TOIS. 1988;6(4):303–331.