

RESEARCH PAPER

RIDAL – A Language for Research Information Definition Argumentation

Mathias Thomas Riechert

German Center for Higher Education Research and Science Studies, Germany
rieichert@dzhw.eu

1 Abstract

Information about the research process is gaining importance for research documentation and evaluation. With the increased usage of such research information, the requirements for data quality and interpretation consistency are increasing. An agreed understanding of the concepts of research information is therefore crucial for fair science evaluation and science policy. Initiatives like euroCRIS and CASRAI address this by standardising research information definitions. In this paper, we present an approach to systematically develop and document not only definitions of research information, but also discussed alternatives and related arguments. With that we aim to support existing RI standardisation initiatives with a flexible and scalable way of documenting and communicating the standardisation process in order to increase acceptance for the resulting definitions. Our contribution is threefold: Based on the widely used IBIS notation for argumentation modelling, we first introduce semantic rules for defining research information. Secondly, a transformation algorithm is provided to reduce the complexity of those argumentations – without the loss of information – and in turn improve readability of the diagrams. Thirdly, the semantic rules of the resulting less complex RIDAL notation are provided. The presented modelling notations are evaluated in the case setting of the standardisation project for research information of the German science system “Core Research Dataset”.

Keywords: research information definition argumentation; research information standardisation; computer-supported argumentation visualisation; issue-based information systems (IBIS); wicked problem

2 Introduction

In recent years, the expectations towards reports about the research process have risen. This is caused in part by new public management, where more organisational freedom comes with a higher duty to report about results. Furthermore, international benchmarking makes efficient processing and documentation of this information necessary. At the same time, the complexity of IT systems and their data structures variability rose, causing an increasing burden of reporting over heterogeneous systems and data for research institutions. In order to reduce the effort required for reporting of research information a growing interest has been seen among researchers, funders, governments and the public in standardising research outputs. By agreeing on a shared understanding of research information definitions, the reporting workload for each researcher can be reduced, as a higher percentage of the reporting information can be obtained from Current Research Information Systems (CRIS) using those standards. Furthermore, the data can be collected, processed, managed and transferred more efficiently, freeing research resources for research rather than research reporting.

For this purpose, information about publications, projects, organisations, persons, products, patents, services, equipment and facilities can be described and managed in the metadata model CERIF (Common European Research Information Format). Documenting such research information (RI) allows for a more

knowledge-driven evaluation of research by providing the possibility of analysing the quality of research datasets and aggregating them across various institutional data sources to ensure access to scientific knowledge (Jörg, 2010) Current Research Information Systems (CRIS) were developed to provide researchers, research managers, innovators, and others with an overview of research activity in a specific domain (Asserson and Jeffery 2010).

As strategic decisions in research institutions are increasingly based on RI the need for higher data quality gets more pressing as well. Especially when RI is used to assess research on a national level (like in national CRIS systems in the Netherlands (Dijk 2012), Norway (Sidselrud and Lingjærde 2012), Slovakia (Turňa et al. 2012) and more listed in the euroCRIS DRIS Directory), a commonly agreed understanding of how the information stored in the CERIF data fields are to be interpreted is crucial for fair evaluation. To illustrate this point: If it is not clear whether or not the definition of *professor* (and therefore the respective data field) includes *honorary professors*, the different stakeholders will document the information inconsistently. This can lead to skewed evaluations, and potentially even abuse by the stakeholders if the information is used as a basis for funding distribution. Another example of problems caused by ambiguous RI definitions is semantic interoperability: The different ways in which standards like CERIF are mapped from one research institution to another (both with varying interpretation of the same RI field) causes interpretation problems when comparing those institutions.

These ambiguity problems can be addressed by bringing all relevant stakeholders to agree on common definitions for RI. CASRAI is an initiative to collaboratively develop agreed definitions of RI based on stakeholder requirements (Baker 2013, Jörg et al. 2014). Building on that approach, we present a language that supports such agreement processes on a large scale by systematically modelling not only the finally agreed-upon definitions, but also the alternatives which have been discussed and respective arguments. By systematically collecting and documenting information about the standardisation argumentation we aim to increase the acceptance of the standardisation result¹, the adoption of research information standards and consequently the efficiency of research reporting. In prior research we presented and implemented a framework on how large-scale definition standardisation processes can be supported by argumentation visualization (Riechert, Biesenbender, & Quix, 2016).

The proposed language is adapted from the Issue-Based Information System (IBIS) notation, which was developed by Kunz and Rittel (1970) to support agreement in complex design processes with multiple stakeholders and conflicting interests. As argumentation models in IBIS get very complex in real-world definition applications (Riechert et al. 2016), we reduce the complexity of IBIS by introducing four modelling rules to allow for better readability and automated post-processing of the definitions documented. The resulting short notation is called RIDAL (Research Information Definition Argumentation Language).

The rest of this article is structured as follows: In Section 2, related work in Computer-Supported Argumentation Visualisation (CSAV) and Design Rationale (DR) is discussed. Section 3 specifies requirements for a language to model RI definitions based on a case study design. The use case we employ is the standardisation project “Specification Project of the German Research Core Dataset” (RCD). RIDAL’s rules and semantics are described in Section 4. Section 5 provides a technical evaluation and Section 6 details our conclusions.

3 Related Work

To the best of our knowledge, no formal approach to modelling RI definitions, including their alternatives and arguments, has yet been discussed in research literature. To introduce formal modelling to our use case, we draw on CSAV literature. We first discuss the nature of the RI definition process as a ‘wicked problem’. Secondly we give an overview of Design Rationale (DR) literature. A design rationale (DR) is a representation of the reasoning behind the design of an artefact (Shum and Hammond 1994). Thirdly, we discuss the most commonly used notation, IBIS, its most important successor – gIBIS – and the most commonly used tool, Compendium.

¹ The link between providing argumentation and the acceptance of the resulting policies is discussed in detail by De Fine Licht, Naurin, Esajasson and Gilljam (2014).

3.1 RI definition as a ‘wicked problem’

Rittel and Webber (1973) define wicked problems as complex design problems “for which no single computational formulation of the problem is sufficient, for which different stakeholders do not even agree on what the problem really is, and for which there are no right or wrong answers, only answers that are better or worse from different points of view” (Introne et al. 2013: 45). Wicked problems are to be distinguished from ‘tame’ problems. ‘Tame’ problems have a well-defined statement. It is clear what they are and what they are caused by. They therefore require a systematic methodology typical of engineering or scientific inquiry, because they belong to a class of problems that can be solved in a similar way each time. Furthermore, they have a definite stopping point and problem solving has no bearing on future options (Jentoft and Chuenpagdee 2009, Rittel and Webber 1973). In contrast to these problems, Rittel and Webber (Rittel and Webber 1973) identify characteristics of a ‘wicked problem’, which may be divided into Robert’s (Roberts 2000) two problem dimensions:

The **problem’s definition**: (1) There is no unique formulation of a wicked problem; (2) Wicked problems do not have a stopping rule; (3) Every wicked problem is unique; (4) Every wicked problem can be regarded as a symptom of another problem; (5) The choice of explanation determines the nature of the problem’s resolution; (6) The multiple stakeholders cause social complexity.

The **problem’s solution**: (7) Solutions to wicked problems are not true-or-false, but good-or-bad; (8) There is no test of a solution to a wicked problem; (9) Since it is impossible to learn by trial-and-error, every solution attempt has irreversible consequences; (10) Wicked problems do not have an enumerable set of potential solutions, nor is there a well-described set of permissible operations; (11) The planner has no right to be wrong.

Preceding qualitative research supports the interpretation of defining RI as a ‘wicked problem’ (Riechert et al. 2016). As argumentation modelling has been specifically designed to address wicked problems, we introduce it as a way to document, model and manage RI definition.

3.2 Modelling argumentations in design processes

In the literature on Computer-Supported Argumentation Visualisation (CSAV), a number of modelling approaches have been developed. In their extensive review, Scheuer et al. (2010) discussed 50 different past and present argumentation systems and 13 empirical studies analysing the effects of their usage. Out of them, the research stream following Rittel’s IBIS directly addresses socially complex or wicked problems, which are IBIS, DRL and QOC. In the following we will briefly discuss their differences and the focus of their application.

The Issue-Based Information Systems (IBIS) notation was developed by Kunz and Rittel (Kunz and Rittel 1970) to support complex design processes with multiple stakeholders with conflicting aims. In IBIS, issues, positions and arguments are set in relation to one another (see Section 3.3 for details). Maps in IBIS notation do not evaluate the strength or weakness of arguments. Direct connections between arguments are explicitly disallowed (Conklin 2005). Therefore, the arguments have to be examined on their own. This is beneficial for depersonalising the discussion (Shum, Selvin, Sierhuis, Conklin, Haley, et al. 2006) but has shown itself to be an issue in contexts like decision argumentation modelling. The Decision Representation Language (DRL) (Lee 1989) addressed this need for documenting decisions in IBIS, and added Goal and Procedure node types and contained a more explicit grammar for relation types. According to Lee, making goals explicit by modular representation enables different viewpoints. Furthermore, arguments are not provided in the form of nodes but as a set of claims.

Design Space Analysis uses a semiformal notation called QOC (Questions, Options, and Criteria) and was developed mainly by MacLean and McKerlie (MacLean et al. 1991, 1993, MacLean and McKerlie 1995, McKerlie and MacLean 1994) to represent Design Rationale. Similarly to IBIS, QOC focuses on basic concepts like Questions and Options. In contrast to IBIS, criteria are used to evaluate and choose an option. Furthermore, DSA emphasises retrospectively rationalising the DR to clarify the dimensions defining the space, as opposed to recording the design process for a single design (Shum and Hammond 1994). Shum stated that both IBIS and QOC evolved into the current Compendium approach and tool (Shum, Selvin, Sierhuis, Conklin, Haley, et al. 2006: 2).

The research stream around IBIS addresses complex design processes with multiple stakeholders. Using DRL instead of IBIS would additionally allow for argument weighting. However, the depersonalisation effect is important for our use case. As such we will focus on IBIS and its current representation in Compendium in the following.

3.3 Issue-Based Information Systems (IBIS), Graphical IBIS (gIBIS) and Compendium

IBIS aims at guiding the identification, structuring, and settling of issues raised by problem-solving groups, and provides information pertinent to the discourse. Although Kunz and Rittel stated that IBIS has the elements topics, issues, questions of fact, positions, arguments, and model problems (Kunz and Rittel 1970), only issues, positions, and arguments were used in later tools (Conklin and Begeman 1988, Shum 2003).

In IBIS, each *Issue* can have many *Positions*. A *Position* is a statement or assertion which responds to the *Issue*. Often *Positions* will be mutually exclusive of each other, but the method does not require this. Each of the *Issue's Positions*, in turn, may have one or more *Arguments* which either support that *Position* or object to it. Thus each separate *Issue* is the root of a (possibly empty) tree, with the children of the *Issue* being *Positions* and the children of the *Positions* being *Arguments* (Conklin and Begeman 1988). The relationships between the *Issues* are shown in **Figure 1**.

Later tools like “Graphical IBIS” (gIBIS) (Conklin and Begeman 1988) implemented the proposed elements of IBIS and provided a graphical interface for modelling the argumentation process. The tool is “designed to support the collaborative construction of these networks by any number of cooperating team members spread across a local area network” (Conklin and Begeman 1988). The aim of gIBIS was to address the interface problems inherent in capturing large amounts of informal design information and in providing effective methods for indexing and retrieval within that information (Conklin and Begeman 1988). **Table 1** lists the properties of IBIS elements and their relationships as stated by Kunz and Rittel (1970) and Conklin and Begeman (1988).

The gIBIS java software was licensed by the Open University’s Knowledge Media Institute to further develop and release the software application and code. Called Compendium (Selvin et al. 2001), this tool was developed to provide an open environment for IBIS modelling, which can be integrated into design workflows. In 2006, Compendium was used on over 100 projects. We refer to the overview by Shum et. al. for empirical evidence of the approach’s learnability and effectiveness (Shum, Selvin, Sierhuis, Conklin, Haley, et al. 2006). **Figure 2** shows the basic IBIS node types as used in Compendium (Shum, Selvin, Sierhuis, Conklin, Rowley, et al. 2006). Compendium provides the concept mapping functionality of gIBIS (implemented in the tool Questmap) but adds a more intuitive user interface and tagging functionality, a sub-map node type, an open architecture as well as API access (Shum, Selvin, Sierhuis, Conklin, Haley, et al. 2006).

The main visual representation in Compendium is a hierarchy of maps. Starting from the base map, it is possible to model multiple facets or hierarchy levels of the present model content. The exporting function to web maps allows the generation of interactive web maps of the Compendium maps. Additionally, the maps can be exported as XML, JPEG, or in a summarised outline form.

In our use case of documenting RI definitions (see Section 4) we modelled over 300 definitions with over 600 arguments. Consequently, the resulting network diagrams became very difficult to navigate and understand. We therefore discuss a simplified adaption of IBIS in Sections 5 and 6.

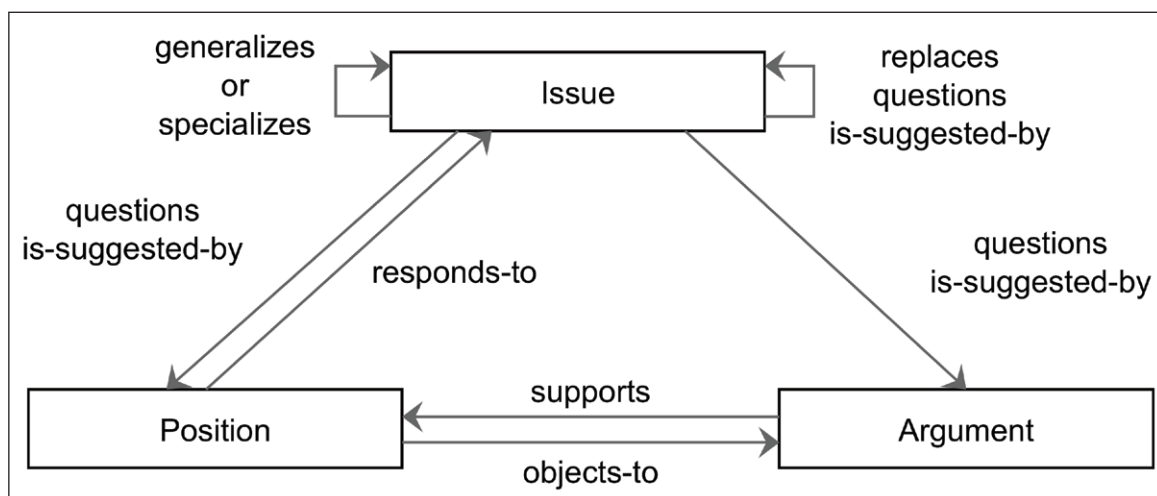


Figure 1: Relationships between IBIS elements (Conklin and Begeman 1988).

Element	Properties
Issues	<p>Issues are the organisational “atoms” of IBIS-type systems:</p> <ul style="list-style-type: none"> • Issues have the form of questions. • The origins of issues are controversial statements. • Issues are specific to particular situations; Positions are developed by utilising particular information from the problem environment and from other cases claimed to be similar. • Issues are raised, argued, settled, “dodged,” or substituted.”
Positions	<p>A Position is a statement or assertion which resolves the Issue. A logically closed set of possible Positions or an open list of possible Positions may be assigned to each issue.</p>
Arguments	<p>Arguments are constructed in defence of or against the different Positions until the Issue is settled by convincing the opponents or decided by a formal decision procedure.</p>
Relationships	<p>There are several kinds of Relationships between Issues, forming networks between the Issues which can be used to aid the search for similar Issues, the history of an Issue, the consequences of previous decisions, etc.:</p> <ul style="list-style-type: none"> • Issue I2 is a direct successor of Issue I1: I2 challenges a statement made in support of one of the Positions maintained in view of I1. • Issue I2 is a generalisation of I1. • I2 is a relevant analogy to I1: the Arguments used in I2 are transferred into Arguments regarding I1, mutatis mutandis. • Positions taken in response to I1 can be compatible, consistent, or incompatible with a Position assumed in response to I2 (by the same or another proponent).

Table 1: Properties of IBIS elements.

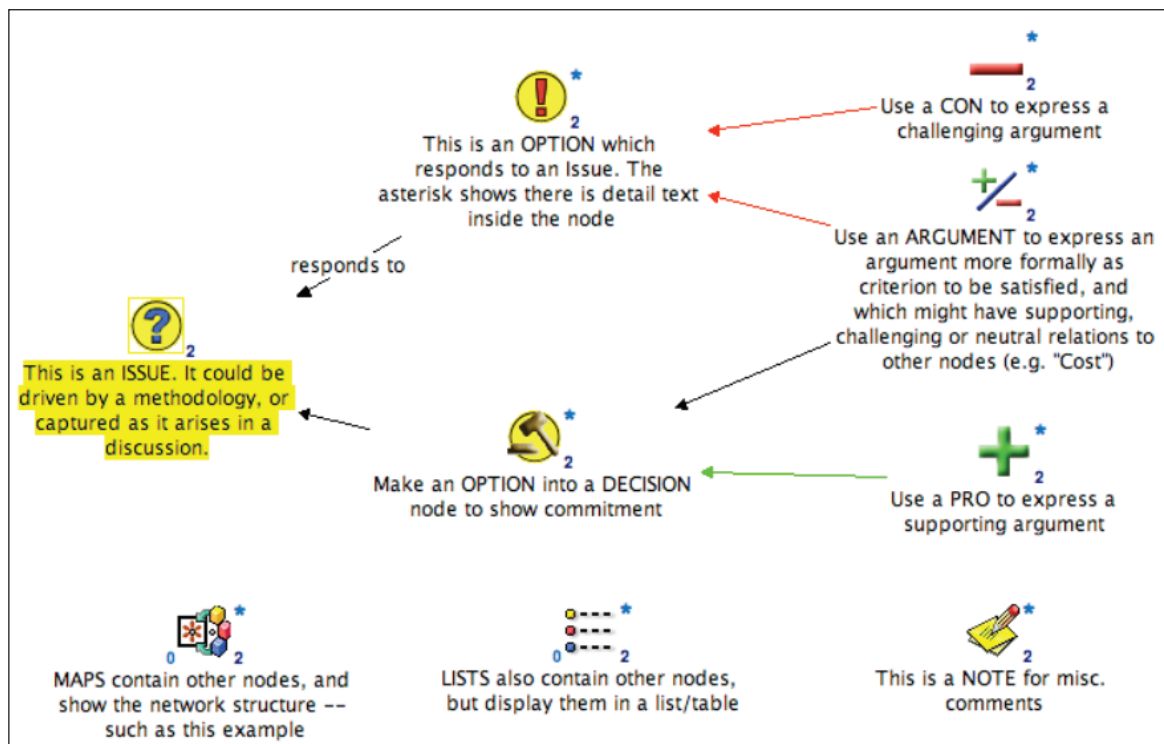


Figure 2: IBIS nodes in Compendium (Shum, Selvin, Sierhuis, Conklin, Rowley, et al. 2006).

4 Requirements

In Requirements Engineering, FURPS (Functionality, Usability, Reliability, Performance, Supportability) is used to support consistent and coherent software requirement definition. These criteria were extended by Grady (1992) to include the four requirement layers Design, Implementation, Interface and Physical Realisation into FURPS+. Considering that our proposed formal language is not a full software system, we focus on the requirement categories functionality and usability. We draw the requirements from three sources: (1) Our own experience in moderating a discussion process in Compendium, modelling the argumentation and developing tools for documentation based on IBIS XML; (2) Feedback we got from the over 50 definition experts involved in the discussion process; (3) Feedback from the research information standardisation community (specifically, at the International Conference on Current Research Information Systems and the euroCRIS initiative). The following IBIS refinement requirements were identified.

Functionality

Added value (R1): The structures IBIS provides form a valuable basis for modelling and visualising the argumentation of definitions. For modelling definitions, the general construct issue is specialised into three types of sub-issue. By using formalised sub-questions, information representation efficiency can be improved, because the resulting network structure follows semantic rules which allow for leaving out elements (as described in Section 5). For this definition case, the sub-questions for attributes and categories of differentiation are addressed:

- Possible characteristics of a defined element can be made explicit. For example, the definition for the element “gender” only has three characteristics: “female”, “male” or “n/a”.
- The second question we constantly face concerns the differentiation of the definition elements: For example, when defining “staff”, differentiation among the categories “full-time equivalents” and “head count” can be defined. They again could be differentiated among the categories nationality, gender, funding, qualification, etc.

Interoperability (R2): The formal language needs to be compatible with the IBIS notation to allow for existing argumentation to be reused. This requires a transformation mechanism to convert IBIS to RIDAL and back.

Usability

The present IBIS notation seems to be a valuable basis for modelling definition argumentation. Additionally, there is excellent tool support with Compendium. However, the resulting maps become very complex when defining research information with their attributes and differentiations. For higher usability, we propose semantic rules to reduce the complexity of argumentation maps in IBIS notation. The agreement on a shared static semantic allows for a comparably simple formal language which forms the basis of application in multiple use cases and tools. A simplification is needed in two regards:

Objective usability (R3): At present there is no format in which argumentation maps can be exchanged without the need for a considerable amount of additional information. While Compendium allows for XML and Web Map export, the resulting files contain a lot of additional information. A formal language should reduce this complexity in terms of the space used and depth of the data structure tree required to describe the same contents.

Subjective usability (R4): Based on experience within the project and discussions with the research information standardisation community, the requirement for better data accessibility became evident. The present visual representation in Compendium is easy to use and intuitive for small maps. In the definition context of our project, with about 80 hours of discussion time the maps became too complex to be printed or even viewed on a screen, even when using sub-maps. We assert that when it comes to representing large-scale discussions, IBIS maps are not accessible enough for external documentation. The current export functionalities (XML, Web Maps) provide direct access to the underlying structure, but are too complex for printing or providing insight. A formal language that incorporates a shared static semantic has to be more easily understood by experts.

Table 2 provides an overview of the requirements addressed.

Requirement	Sub-requirement	Description
Functionality	R1: Added value	Explicit visual and formal representation for the underlying questions for attributes and categories of differentiation.
	R2: Interoperability	Transformation from and to IBIS
Usability	R3: Objective usability	Reduce complexity in terms of space required
	R4: Subjective usability	Reduce visual and representational complexity of the argumentation diagrams

Table 2: Definition of Argumentation Language requirements.

5 Language Specification

This section presents the specification of RIDAL to meet the requirements described in Section 4. To begin with, the use case is introduced as an example. Section 5.2 discusses the modelling of definitions of research information in pure IBIS. Section 5.3 introduces semantic rules for modelling RI in IBIS in a more structured way. Following these rules allows for automatic conversion of the definition model and reuse in other applications. Section 5.4 explains how definition models following those rules can be transformed into a less complex representation without any loss of information. Section 5.5 presents the rules for modelling directly in the less complex RIDAL notation. These three forms – pure IBIS, rule-based IBIS and RIDAL – are the subject of evaluation in Section 6.

5.1 Case “Specification Project of the German Research Core Dataset”

The use case we employ is the standardisation project “Research Core Dataset” (RCD) for the German science system. The project was initiated in 2013 by the German Council of Science and Humanities to develop a shared set of definitions for research information (e.g. staff, publications, funding, patents) for the German science system. More than 48 different stakeholders are involved in the project. They include representatives of universities, non-university research institutions, ministries, research information system vendors and scientific societies. Furthermore, four pilot universities are integrated in the project. The definition process is organised in four project groups, each of which has eight experts. Each of the project groups holds up to six meetings, with 1–2 days of discussion time per meeting (with eight hours of discussion time per day). The project group “definitions and data formats” defines research information for all six research information areas. The discussions are structured and moderated by using CSAV on a central screen. After an initial discussion phase, a feedback phase is conducted. All pilot universities, the non-university research institutions and software vendors are asked for feedback concerning the applicability of the definitions. After the feedback round, another discussion phase is conducted to integrate the various external feedback into the definition specification.

This case allows us to analyse the value of argumentation modelling for reaching a definition consensus in a particularly complicated case (or ‘extreme’ case; see Seawright and Gerring 2008). The German case exemplifies extreme characteristics because the country’s federal regulation of higher education and research institutes entails particularly fragmented and diverse reporting processes and requirements (Biesenbender and Hornbostel 2016), resulting in particularly complex stakeholder positions. Testing the language in this complex case will result in higher applicability in less complex cases.

5.2 Modelling the definition of research information in IBIS

The IBIS notation as proposed by Kunz and Rittel had no explicit formal model (Kunz and Rittel 1970), but was a paper-based approach. The implicit notation was later implemented in gIBIS (Conklin and Begeman 1988). Conklin and Begeman provided the first semi-formal depiction of the elements and their relationships (see **Figure 1** in chapter 3.2). The most recent successor is Compendium.

Section 3.3 introduced the modelling of IBIS in Compendium. All IBIS node types have a representation in Compendium. Issues are modelled as questions (and displayed as a node with a question mark). Positions are modelled as options (and displayed with an exclamation mark). Arguments are displayed as nodes with a plus and minus sign. Compendium additionally allows for sub-maps, encapsulating more detailed discussion content. In the following we will stick to IBIS terminology (issues and positions). Note that the Compendium terms (questions and options) could be used correspondingly.

When modelling definitions of research information in Compendium, all elements can be placed freely on a canvas. Furthermore, there are no restrictions about which kind of issues can be modelled. This offers a high degree of flexibility for a wide range of IBIS applications, but results in very complex and confusing diagrams. The possibility of using sub-maps allows for encapsulation, but is not very user-friendly if the number of sub-levels rises above three. In our case, we had up to 9 sub-levels, which is very challenging to keep track of for all persons trying to read the argumentation model of the RI definitions.

In order to be able to simplify the definition model without losing any information and to meet the usability requirements R3 and R4, in the next section we discuss the introduction of semantic modelling rules.

5.3 Introducing rules for definition modelling of research information in IBIS

Using the IBIS notation in Compendium, a sample definition for “full-time equivalent” could be modelled as shown in **Figure 3**. The root question node (=issue node in IBIS) asks for a suitable definition for the element “full-time equivalent” in the research information context. The definition itself is documented in the node. Its definition can therefore be shown by clicking on the node in Compendium. Two possible alternative options (equivalent to position nodes in IBIS) have been discussed in the definition process by the experts. Both alternative options have arguments supporting or challenging the option.

On the next level, each of the alternative options can have the sub-question “How can . . . be differentiated?”. Alternative options can be appended to this question on the next level. If the differentiations of a definition are an exhaustive list, attributes should be used instead of differentiations. This is represented by the question “What are possible attributes for . . .?”. Using this approach allows for freely scalable maps to model as detailed differentiations as necessary, with the option of providing and documenting arguments on each level of the definitions. It has to be noted that IBIS allows for any type of question to be modelled. By restricting the usage of IBIS questions (or issues) to the above types of question, a higher degree of formalisation can be achieved. This forms the basis of the simplification described in Section 5.4.

We formalise the semantic rules described for modelling definitions in IBIS:

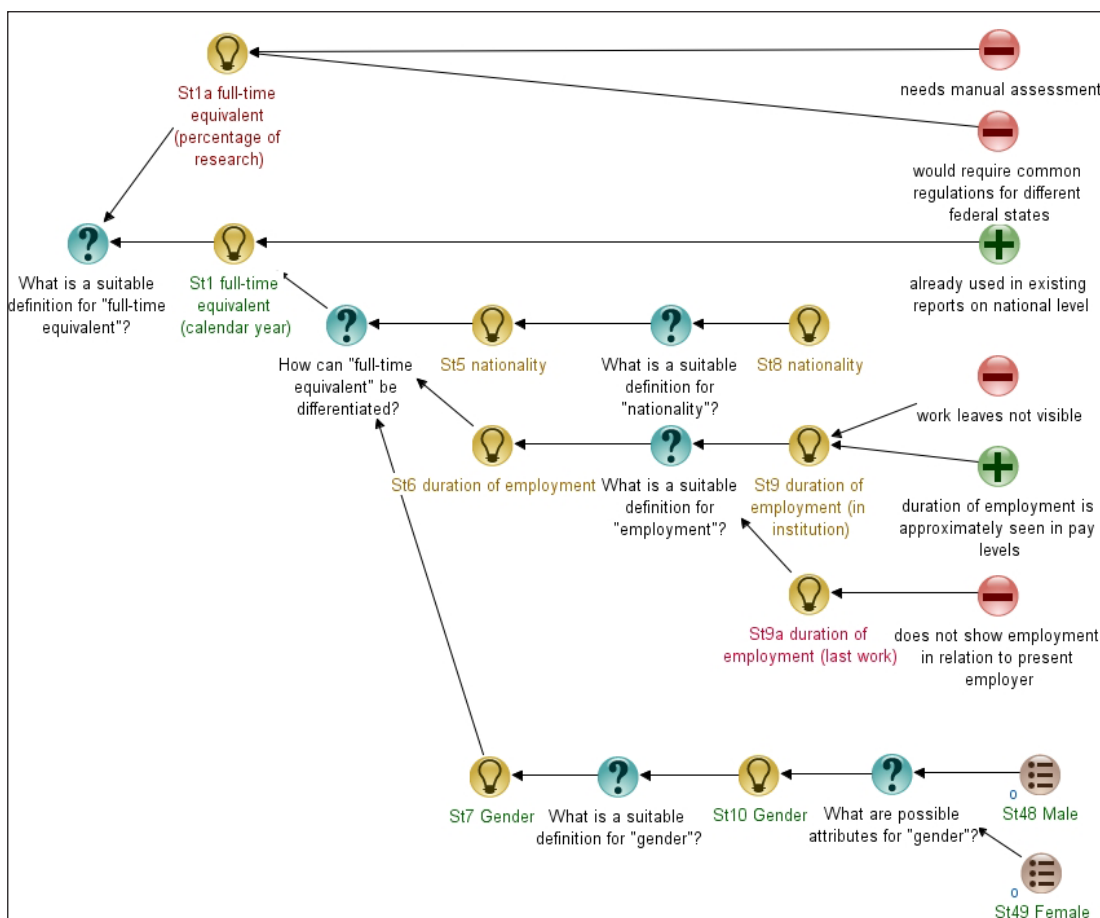


Figure 3: Excerpt defining “full-time equivalent” in Compendium using the IBIS notation.

- **Node rule 1: ID usage in the definition label:** In order to allow for fast searching and identification, a serial ID is added to the position nodes.

In the example: “St” shows that the definition deals with staff. The number is a serial number for all elements defined. Alternative positions (options in Compendium) like St1 or St1a are appended with a,b,c,d, . . . for each alternative position.

- **Node rule 2: Issue types:** Issue nodes can be one of three types:

- Type A: Definition Issue: “What is a suitable definition for . . .?”
- Type B: Differentiation Issue: “How can . . . be differentiated?”
- Type C: Attribute Issue: “What are possible attributes for . . .?”

Note: In the following, positions addressing those issues are called type a (for addressing definition issues type A), type b (for addressing differentiation issues type B) and type c (for addressing attribute issues type C).

- **Node rule 3: Issue type A (Definition) as tree root:** Each definition tree starts with the issue (or question in Compendium) “What is a suitable definition for . . .?”.

In the example: “What is a suitable definition for full-time equivalent” is the root node.

- **Link rule 1: Position answers issue:** Positions (or options in Compendium) addressing that issue are connected to that issue. The definition text of the element is provided in that position node.

In the example: Two alternative options are provided as positions (St1 and St1a).

- **Link rule 2: Differentiation and attributes as sub-issues of positions:** Each position can be further defined by a sub-issue asking for differentiation (Node rule 3) or for attributes (Node rule 4). If no more detailed definition is required, stop here.

In the example: “St1 full-time equivalent (calendar year)” is connected to a differentiation issue, and “St7 Gender” is connected to an attribute issue.

- **Node rule 4: Sub-issue type B (Differentiation):** “How can . . . be differentiated?”. Positions for that issue are added based on Link rule 1.

- **Node rule 5: Sub-issue type C (Attributes):** “What are possible attributes for . . .?” Positions for that issue are added based on Link rule 1.

- **Node rule 6: Attributes as different node type:** Positions for attributes receive a different node type. This can be achieved by tagging and using a different visual representation.

In the example: “St48 Male” has a different node icon than “St10 Gender”, because it is an attribute and not a differentiation.

- **Link rule 3: Restart from Node rule 2 for differentiation or attribute definition:** Each differentiation or attribute position is defined by adding another root issue tree (i.e. by restarting the rules from Node rule 2 onwards).

In the example: “St5 Nationality” is followed by another Root issue “What is a suitable definition for Nationality”. As the appended position (including the definition text) is detailed enough, no further differentiation or attributes are defined below.

- **Link rule 4: Arguments support and challenge positions:** Pro- and con- arguments can be added to positions (not to questions!) on every level.

- **Node rule 7: Set usage status as tag:** For each position (or option) it is decided whether the definition is to be used or not used (for example for alternative positions not being used). The relevant information can be tagged as “used”, “not used”, and “more definition work required” in Compendium. Additionally, this information can be visually represented by using text colours green, red and orange.

In the example: “St1” is tagged as “used”, while its alternative position “St1a” is tagged as “not used”. This is visually represented by green and red text colour.

Additionally, we use placement rules for higher readability. These are only required if the diagrams are to be read without further transformation.

- **Placement rule 1: Nodes on the same depth level at same x position:** To improve readability of the diagram, all nodes on the same depth level are placed at the same x position.

In the example: Both alternative positions of full-time equivalent St1 and St1a are on the same depth level and therefore placed at the same x position.

- **Placement rule 2: Sub-issues and positions from left to right:** To improve readability of the diagram, all sub-issues and positions are placed from left to right.
In the example: All positions and sub-issues are placed to the right of the issue.
- **Placement rule 3: Arguments on the right border of the diagram:** To improve readability, all arguments are placed on the right side of the diagram.

Fehler! Verweisquelle konnte nicht gefunden werden. shows an overview of how to model in Compendium using the above rules.

5.4 Model transformation to RIDAL

Based on these rules, it is possible to relate the different positions (or options), differentiations and attributes without having to specify the underlying issues (or questions) as shown in **Figure 4**. By using the conversion rules stated below, an IBIS diagram modelled using the rules in Section 5.3 can be converted to RIDAL without any loss of information.

For each position type b (i.e. addressing a differentiation issue) and type c (addressing a attribute issue):

- Connect all linked argument nodes to the next position type a (addressing definition issue) two levels lower.
- Remove connection to the position type b.

For each position type a (i.e. addressing a definition issue):

- Connect all positions on the same level with an arrowless link with the label “Alternative” (*see step 1 in Figure 5*).
- Get all child positions type A (definition) four levels lower, remove all connections it has to its direct parent elements (issue type A) and connect it to the parent position type A (four levels higher). A special rule applies for reconnection: Only reconnect the positions if their status is not “not used”. If all of the alternative positions have the “not used status”, connect the position with the smallest y position anyway (so there is always at least one connected element) (*see steps 2 and 3 in Figure 5*).
- Remove all child issues type B (differentiation) and C (attributes) as well as their related links (*see step 4 in Figure 5*).

Finally:

- Remove the root issue type A (definition)

The resulting diagram is shown in Section 5.5 in **Figure 6**. Note that the transformation can be reversed by reversing the order of the steps and inverting them (i.e. “create” instead of “remove”, “remove” instead of “connect”).

5.5 RIDAL

By allowing only three question types and following the rules in Section 5.3, a much simpler representation of the content is possible without any loss of information by using the transformation specified in Section 5.4. The resulting model is shown in **Figure 6**.

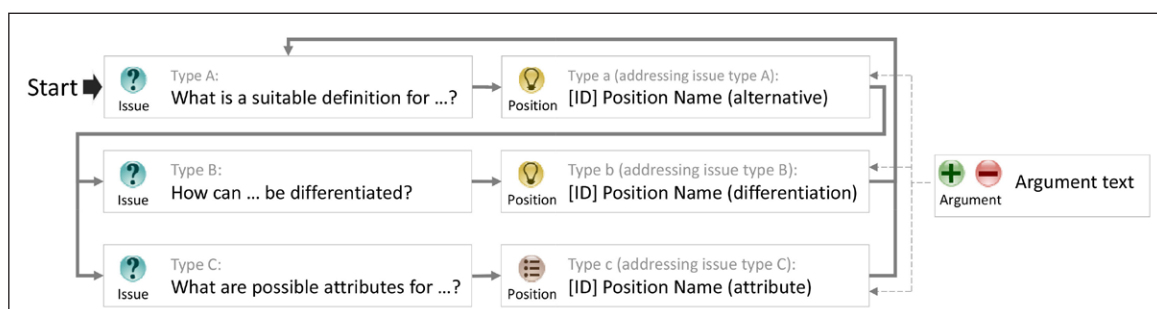


Figure 4: Modelling research information definitions with IBIS.

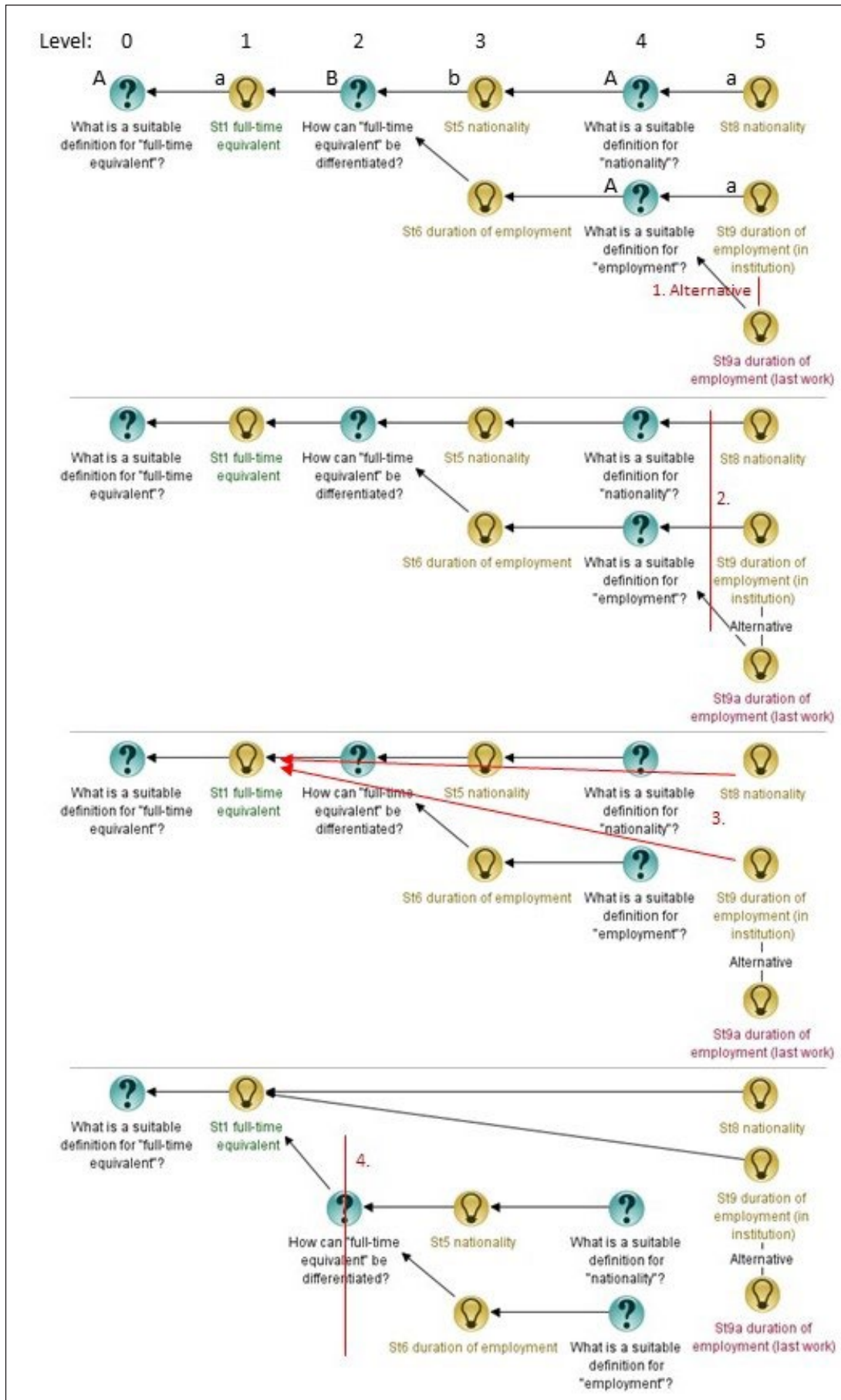


Figure 5: Main steps of transforming rule-based IBIS to RIDAL.

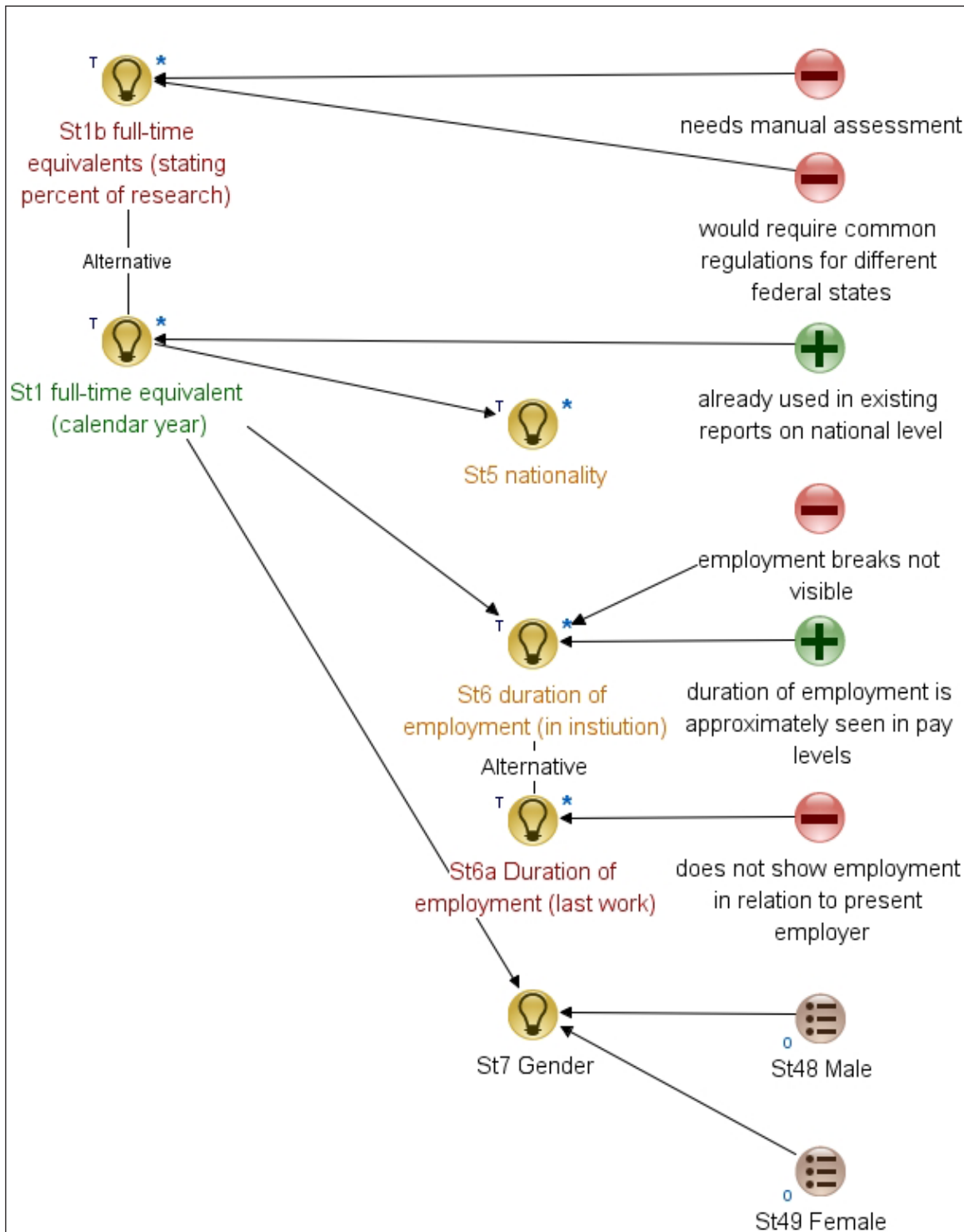


Figure 6: Excerpt defining “full-time equivalent” in Compendium using RIDAL.

In order to model directly in RIDAL, the rules from our IBIS modelling in Section 5.3 can be adapted:

- **Node rule 1: ID usage in the definition label:** All positions (or options in Compendium) are connected to that issue. The definition text of the element is provided in that position node. *In the example: Two alternative options are provided as positions (St1 and St1a).*
- **Node rule 2: Position as tree root:** Each definition tree starts with the position (or answer in Compendium) for the implicit question “What is a suitable definition for . . .?”. If there is more than one possible definition, each gets its own position node and they are connected with an arrowless link with the label “Alternative”.

In the example: Position “St1 full-time equivalent (calendar year)” is the root node. Its alternative St1a is connected with an “Alternative” link.

- **Link rule 1: Differentiation and attributes as sub-positions:** Each position can be further defined by a sub-position asking for differentiation (Node rule 3) or for attributes (Node rule 4). If no more detailed definition is required, stop here.

In the example: “St1 full-time equivalent (calendar year)” is connected to the differentiation positions St5, St5 and St7. “St7 Gender” is connected to the attribute positions St48 and St49.

- **Node rule 3: Sub-position addressing differentiation:** Definition text answering the implicit question “How can . . . be differentiated?”
- **Node rule 4: Sub-issue addressing attributes:** Definition text answering the implicit question “What are possible attributes for . . .?”
- **Link rule 2: Restart from Link rule 1 for differentiation or attribute definition:** Each differentiation or attribute is defined by adding links for differentiations or attributes below (i.e. by restarting the rules from Link rule 2 on).

In the example: “St7 gender” is further defined by attribute positions below.

- **Link rule 4: Arguments support and challenge positions:** Pro- and con- arguments can be added to positions (not to questions!) on every level.
 - **Node rule 5: Set usage status as tag:** For each position (or option) it is decided whether the definition is to be used or not used (for example for alternative positions not being used). The relevant information can be tagged as “used”, “not used”, and “more definition work required” in Compendium. Additionally this information can be visually represented by using text colours green, red and orange.
- In the example: “St1” is tagged as “used”, while its alternative position “St1a” is tagged as “not used”. This is visually represented by green and red text colour.

Figure 7 shows an overview of how to model research information definitions in RIDAL using the above rules. Note that in contrast to the IBIS, position nodes type a (alternatives) have to be linked to each other with the label “Alternative” to be able to distinguish between the position types.

6 Evaluation

To assess the proposed language’s value, we structure its evaluation according to the requirements defined in Section 4.

6.1 Requirement 1: Added value

The central idea of language simplification is to use only three types of question: “What is a suitable definition for . . .”, “How can . . . be differentiated?”, and “What are possible attributes for . . .?”. By making them explicit, experts defining research information have a structure to guide them through the discussion. This structure also helps to ensure that each definition is checked if further differentiation is required and if alternative definitions might be better suited. This increased formalisation comes at the cost of less flexibility when other question aspects might be of interest. We consider this requirement to be met.

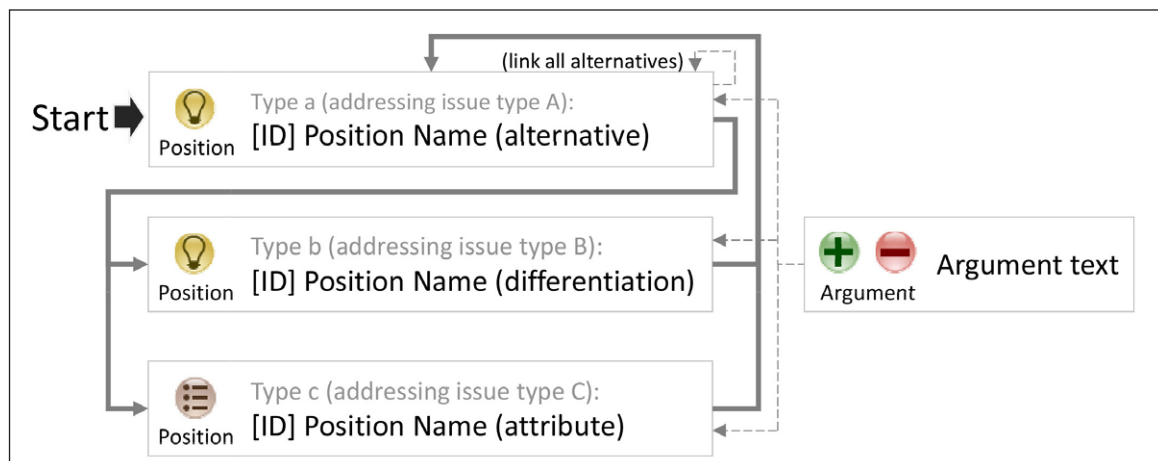


Figure 7: Modelling research information definitions with RIDAL.

6.2 Requirement 2: Interoperability

The semantic rules for modelling definitions in IBIS and RIDAL described in Section 5 allow for conversion of IBIS definitions to RIDAL definitions and vice versa without any loss of information (see Section 5.4). By using the steps as a conversion algorithm, definition models in IBIS can be converted automatically to RIDAL and back. Therefore it is possible to use both representation forms depending on usage context. For example, it might be of benefit to use the shorter RIDAL in situations where modelling speed is of importance, for instance when modelling is performed on-screen while experts discuss the definitions in order to increase model completeness and discuss group agreement (as found in Selvin et al. 2001). By leaving out redundant “What is a suitable definition for . . .” questions but still being able to provide the full definition context based on the semantics, improves the experience for experts in the discussion group and enhances modelling speed. In other cases, having the redundant questions inside the model might be of benefit. For example, when using the diagrams to document what has been discussed, seeing questions might help external experts without knowledge of the semantic rules. Since conversion from and to RIDAL has been shown to be possible, we consider this requirement to be met.

6.3 Requirement 3: Objective usability

In our use case of defining a research information standard for the German science system, experts agreed on 305 definitions (including differentiations and attributes), and discussed 172 alternative definitions. Additionally, more than 600 arguments were documented. By using RIDAL, which is less complex, the resulting text-only XML files have a size of 4 megabytes. When using the Compendium Web Map export (with a visual interface), the total file size amounts to 26 megabytes.

When using the structured IBIS, the text-only XML export has a file size of 8.2 megabytes and the Web Map export has a file size of 50.8 megabytes. Therefore, documenting the same definition content in RIDAL requires 49.4 % of rule-based IBIS for text-only export, and 52.2 % for Web Map export with a visual interface. One might argue that storing files of such sizes is no longer a problem. But when it comes to presenting Web Maps online, this reduction of about 50 % does play a role in mobile applications or if internet access is limited.

As a second dimension of complexity we counted the maximum depth of argumentation tree (i.e. the number of definition sublevels). While the RIDAL has a maximum depth of 14 levels, the corresponding structured IBIS requires 28 levels. RIDAL therefore reduces the tree depth by 50%.

6.4 Requirement 4: Subjective usability

Much more important than the necessary space is the reduction in visual complexity. By using RIDAL instead of a rule-based IBIS notation, in our case the number of definition-related nodes is reduced by 38.4 % from 775 to 447. As the number of arguments stays the same, the total number of nodes required is reduced by 26.3 % from 1405 to 1107.

The second component of visual complexity in network diagrams is the number of links required by the visual representation. Using RIDAL instead of a rule-based IBIS notation reduced the number of links in our project by 55.2 %, from 3085 to 1383.

6.5 Evaluation summary

The evaluation shows that using RIDAL reduces the number of required nodes by 26.3 %, the number of required links by 55.2 %, the text-only space by 49.4 %, the necessary Web Map space by 52.2 %, and the tree depth by 50% without any loss of information. One limitation of using RIDAL is that the readers of the diagrams have to be familiar with the rules. On the other hand, having said rules explicitly available helps when modelling RI definitions, because the process of definition is more structured, thus resulting in higher data consistency compared to modelling completely freely in IBIS. By providing a transformation algorithm, it is possible to convert RIDAL back to rule-based IBIS when required. Finally, the introduction of rules allows for the automatic reuse of the documented information for transforming definitions into a data model or providing additional visual representations to address weaknesses of Compendium Web Maps.

7 Limitations

The application of definition modelling is limited by two main factors. Firstly, modelling not only definitions but also alternatives and arguments requires effort and time. In our project, a group of three people was responsible for the preparation, moderation, on-screen modelling and post-processing of the definition

process for the whole project time of two years. Secondly, providing transparency with regard to discussed alternatives and arguments does also require a certain willingness to be open about the contents discussed. In our case, the transparency provided was received positively by the broad range of participating stakeholders from the science system. However, the expert legitimation is largely replaced by argument legitimation, which places high requirements on the quality of those arguments.

Comparing modelling with the introduced rules to free definition modelling in IBIS, a further limitation is the restricted flexibility. Restricting the modelling to three question types allows for automatic conversion and model reduction, but comes at the cost of a more structured discussion process. As discussed, this also offers benefits for the definition process itself.

Comparing modelling in rule-based IBIS to the presented RIDAL, a lot of redundant nodes are left out to achieve the reduction in complexity. This limits the reading of RIDAL to persons who are familiar with the rules. We address this limitation by providing a modelling overview (**Figure 7**) to assist readers and modelers. Additionally, the models can be converted by using the transformation algorithm in order to get the more intuitive representation in IBIS or the less complex form in RIDAL, depending on the required context.

Another limitation results from our use case approach. Our results need to be tested in other contexts and definition projects on a larger scale in order to further evaluate the usability of RIDAL.

8 Conclusion

Systematically defining research information is an important step towards achieving a commonly agreed understanding of what is meant by talking about information about research. The presented approach employs the central solution strategy from wicked problem literature on RI definition, supporting transparency and acceptance of discussion results by modelling and documenting issues, alternative positions and related arguments. The introduced rules for modelling definitions of research information in IBIS allow for a more structured definition process, conversion of structured IBIS to the less complex RIDAL notation and an export and automatic reuse in other representation forms. By using RIDAL for modelling, the definition model size was reduced by 50 %, and diagram complexity by 40 %, without any loss of information. The structured modelling rules presented for rule-based IBIS and RIDAL allow for both automatic and manual modelling of definition processes. Additionally, a modelling overview (**Figure 4** and **Figure 7**) for both modelling notations is provided to guide practical use of the rules in other research information definition projects. By using a more formalised means of defining research information, initiatives like euroCRIS and CASRAI, and the global research community, could contribute to a systematic knowledge base of research information definitions and arguments.

Further research could focus on subjective usability in order to gain a deeper understanding of the intuitiveness and readability of the rule-based IBIS and RIDAL diagrams. So far, user studies have found IBIS models to be intuitive for external stakeholders (Loukis et al 2009). In our case, which involved a complex IBIS diagram, qualitative feedback demonstrated only partial support for the documentation in Compendium Web Maps. A deeper understanding of the perception and usage of large diagrams in this context might be of high value for the modelling of RI definitions.

Based on the structured definitions and the transformation algorithm, more interactive forms of visual representation of the discussion contents can be developed. Our experience within the project has shown that external insight can only partially be reached by providing Compendium Web Maps. The most central problem was reported to be the high complexity of the diagrams. By applying best practices from information visualisation, we will develop and examine different interaction and representation forms. These newly developed representation forms will then be evaluated in comparison with Compendium Web Maps and protocols to analyse how different forms of this visual representation influence the perceived transparency of the definition process for external stakeholders. This is an important step to examine how far and which dimensions providing information about alternatives and arguments support agreement among different stakeholders.

Another branch of future research could address the ideal degree of detail for the definitions and their arguments. As the proposed approach allows for endless scalability, the resulting model and diagram can become very complex, reducing overall readability. Introducing different degrees of detail for different target groups could potentially address this issue.

Employing CSAV with RIDAL was highly valuable in our project. Applying this structured approach to other standardisation initiatives for research information might help to get more stakeholders to understand the concepts behind, and agree on, common definitions of research information. We believe this to be an important step on the way to a common understanding of research information, improving overall quality and usability of information about research.

Acknowledgements

This work was funded by the German Federal Ministry of Education and Research.

Competing Interests

The author has no competing interests to declare.

References


- Asserson, A** and **Jeffery, K** 2010 CRIS and institutional repositories. *Data Science Journal*, 9:CRIS14–CRIS23.
- Baker, D** 2013 Solving data disconnect in global research administration: the CASRAI approach. *Research Global*, November.
- Biesenbender, S** and **Hornbostel, S** 2016 The Research Core Dataset for the German science system: challenges, processes and principles of a contested standardization project. *Scientometrics*, 106(2): 837–847. DOI: <https://doi.org/10.1007/s11192-015-1816-y>
- Conklin, J** 2005 *Dialogue mapping: Building shared understanding of wicked problems*. John Wiley & Sons, Inc. Abgerufen von: <http://dl.acm.org/citation.cfm?id=1237968>.
- Conklin, J** and **Begeman, M L** 1988 gIBIS: a hypertext tool for exploratory policy discussion. *ACM Transactions on Information Systems (TOIS)*, 6(4): 303–331. DOI: <https://doi.org/10.1145/58566.59297>
- De Fine Licht, J**, **Naurin, D**, **Esaiasson, P** and **Gilljam, M** 2014 When Does Transparency Generate Legitimacy? Experimenting on a Context-Bound Relationship. *Governance*, 27(1): 111–134. DOI: <https://doi.org/10.1111/gove.12021>
- Dijk, E** 2012 Visualization of NARCIS: a network of experts and knowledge organization in the Netherlands.
- Grady, R B** 1992 *Practical software metrics for project management and process improvement*. Prentice-Hall, Inc. Abgerufen von: <http://dl.acm.org/citation.cfm?id=140207>.
- Introne, J**, **Laubacher, R**, **Olson, G** and **Malone, T** 2013 Solving Wicked Social Problems with Socio-computational Systems. *KI-Künstliche Intelligenz*, 27(1): 45–52. DOI: <https://doi.org/10.1007/s13218-012-0231-2>
- Jentoft, S** and **Chuenpagdee, R** 2009 Fisheries and coastal governance as a wicked problem. *Marine Policy*, 33(4): 553–560. DOI: <https://doi.org/10.1016/j.marpol.2008.12.002>
- Jörg, B** 2010 CERIF: The common European research information format model. *Data Science Journal*, 9: CRIS24–CRIS31.
- Jörg, B**, **Höllrigl, T** and **Baker, D** 2014 Harmonising and Formalising Research Administration Profiles CASRAI / CERIF. *Procedia Computer Science*, 33: 95–102. DOI: <https://doi.org/10.1016/j.procs.2014.06.016>
- Kunz, W** and **Rittel, H W** 1970 *Issues as elements of information systems* (Bd. 131). Oakland: Institute of Urban and Regional Development, University of California.
- Lee, J** 1989 Decision representation language (DRL) and its support environment. Abgerufen von: <http://dspace.mit.edu/handle/1721.1/41499>.
- Loukis, E**, **Xenakis, A** and **Tseperli, N** 2009 Using argument visualization to enhance e-participation in the legislation formation process. In: *Electronic Participation*. Springer, pp. S. 125–138. Abgerufen von: http://link.springer.com/chapter/10.1007/978-3-642-03781-8_12.
- MacLean, A**, **Bellotti, V** and **Shum, S** 1993 Developing the design space with design space analysis. *NORTH HOLLAND STUDIES IN TELECOMMUNICATION*: 197–197.
- MacLean, A** and **McKerlie, D** 1995 Design space analysis and use-representations. *Scenario-based design: envisioning work and technology in system development*, pp. 183–207.
- MacLean, A**, **Young, R M**, **Bellotti, V M** and **Moran, T P** 1991 Questions, options, and criteria: Elements of design space analysis. *Human–computer interaction*, 6(3–4): 201–250. DOI: https://doi.org/10.1207/s15327051hci0603&4_2
- McKerlie, D** and **MacLean, A** 1994 Reasoning with design rationale: practical experience with design space analysis. *Design Studies*, 15(2): 214–226. DOI: [https://doi.org/10.1016/0142-694X\(94\)90026-4](https://doi.org/10.1016/0142-694X(94)90026-4)
- Riechert, M**, **Biesenbender, S**, **Dees, W** and **Sirtes, D** 2016 Developing Definitions of Research Information Metadata as a Wicked Problem? Characterisation and Solution by Argumentation Visualisation. *Program*, 50(3). DOI: <https://doi.org/10.1108/PROG-01-2015-0009>
- Riechert, M**, **Biesenbender, S** and **Quix, C** 2016 Developing and Standardising Definitions for Research Information: Framework and Methods of Successful Process Documentation. *Procedia Computer Science*. Abgerufen von: <http://dspacecris.eurocris.org/handle/11366/504>.
- Rittel, H W** and **Webber, M M** 1973 Dilemmas in a general theory of planning. *Policy sciences*, 4(2): 155–169. DOI: <https://doi.org/10.1007/BF01405730>

- Roberts, N** 2000 Wicked problems and network approaches to resolution. *International public management review*, 1(1): 1–19.
- Scheuer, O, Loll, F, Pinkwart, N and McLaren, B M** 2010 Computer-supported argumentation: A review of the state of the art. *International Journal of Computer-Supported Collaborative Learning*, 5(1): 43–102. DOI: <https://doi.org/10.1007/s11412-009-9080-x>
- Seawright, J and Gerring, J** 2008 Case selection techniques in case study research a menu of qualitative and quantitative options. *Political Research Quarterly*, 61(2): 294–308. DOI: <https://doi.org/10.1177/1065912907313077>
- Selvin, A, Buckingham Shum, S, Seirhuis, M, Conklin, J, Zimmerman, B, Palus, C, . . . , Motta, E** 2001 Compendium: Making meetings into knowledge events. In: *Proceedings of the 2001 Knowledge Technologies 2001 Conference*. Austin, Texas.
- Shum, S B** 2003 The roots of computer supported argument visualization. In: Kirschner, P A, Shum, S J B and Carr, C S (Hrsg.) *Visualizing argumentation*. London: Springer, pp. S. 3–24. Abgerufen von: http://link.springer.com/chapter/10.1007/978-1-4471-0037-9_1.
- Shum, S B and Hammond, N** 1994 Argumentation-based design rationale: what use at what cost? *International Journal of Human-Computer Studies*, 40(4): 603–652. DOI: <https://doi.org/10.1006/ijhc.1994.1029>
- Shum, S B, Selvin, A, Sierhuis, M, Conklin, J, Haley, C B and Nuseibeh, B** 2006 Hypermedia support for argumentation-based rationale. In: *Rationale management in software engineering*. Springer, pp. S. 111–132. DOI: https://doi.org/10.1007/978-3-540-30998-7_5
- Shum, S B, Selvin, A, Sierhuis, M, Conklin, J, Rowley, A, Juby, B, . . . , Mancini, C** 2006 FROM gIBIS TO MEMETIC Evolving a Research Vision into a Practical Tool. Abgerufen von: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.105.6403>.
- Sidselrud, A and Lingjærde, G C** 2012 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)*, pp. S. 305–312.
- Turňa, J, Noge, J and Zendulková, D** 2012 The system SK CRIS, scientific publications and theses—mirror of Slovak science. Abgerufen von: <http://dspacecris.eurocris.org/handle/11366/111>.

How to cite this article: Riechert, M T 2017 RIDAL – A Language for Research Information Definition Argumentation. *Data Science Journal*, 16: 5, pp.1–17, DOI: <https://doi.org/10.5334/dsj-2017-005>

Submitted: 02 July 2016 **Accepted:** 05 January 2017 **Published:** 23 February 2017

Copyright: © 2017 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

 *Data Science Journal* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS 