

# A Channel Theory based 2-Step Approach to Semantic Alignment in a Complex Environment

**Andreas Bildstein**

University of the West of Scotland/School of Engineering and Computing, Paisley, PA12BE, UK  
Email: b00230731@studentmail.uws.ac.uk, andreas.bildstein@ipa.fraunhofer.de

**Junkang Feng**

University of the West of Scotland/School of Engineering and Computing, Paisley, PA12BE, UK  
Email: junkang.feng@uws.ac.uk

Received: 31 March 2016; Accepted: 22 August 2017; Published: 08 September 2017

**Abstract**—This article shows a novel approach to semantically align two domain contexts in a distributed system based on the theory of Information Flow [1], also known as Channel Theory. In this article, we propose a 2-step approach to cope with the increasing complexity in constructing the channels, when the channel theory is applied in a complex environment, for example in the area of smart manufacturing. We describe why the methods that had been used so far for constructing the channel might not be suitable for such a complex environment and introduce the main components of our approach. Furthermore, we are explaining how these components work together by using an example from the manufacturing area where product specifications have to be aligned with the production capabilities of manufacturing equipment. Within this example, in the first step a high-level description of production steps is mapped to production processes, and in the second step, a detailed description of the production steps in question is mapped to available equipment and tooling that is related to the filtered production processes from step 1.

**Index Terms**—Information flow, semantical alignment, product specifications, equipment capabilities, complex environment.

## I. INTRODUCTION

The theory of Information Flow (IF for short), also known as Channel Theory [1], provides us with the mathematical tools that help us to describe the flow of information within a distributed system. With the help of these tools, we are able to understand that information can flow between the parts of a distributed system. Since then, we have seen several approaches for the application of this theory, e.g. in the work of Kent [2] and its Information Flow Framework [3], or in the work of Kalfoglou and Schorlemmer, who developed IF-Map [4] based on IF to align ontologies semantically. Especially the various work of Kalfoglou and Schorlemmer in the context of the application of IF, e.g. [5], [6], or [7]

prepared the ground for successive researchers who also applied the theory of IF to real world problems.

What we observe in the work of Kalfoglou and Schorlemmer and the researchers who build up their approaches on this work is that they are mainly using examples with a fixed defined set of instances (tokens) and properties (types) that are not expected to change over the time. For example, Xu and Feng [8] show an example with two questionnaires and a small set of questions within those questionnaires that shall be integrated based on the application of the methodology shown by Kalfoglou and Schorlemmer [4]. Also, Yang and Feng [9] show a non-dynamic scenario where two databases containing employees from two merged companies shall be integrated based on the locations of the employees. Similar implementations of the approaches from Kalfoglou and Schorlemmer can be found for example in [10] or [11], whereby Yang [11] found out that the approach from Kalfoglou and Schorlemmer might be error-prone when the application is getting too complex and showed some new approaches to cope with these drawbacks.

We will show how we think that these approaches might have to be enhanced in the application area of smart manufacturing for coping with the higher degree of complexity in this application area compared with the degree of complexity shown in the examples of the relevant literature so far. For this enhancement, we propose a 2-step approach for semantically integrating the contexts of product specifications with the contexts of production capabilities to reduce the complexity within both the contexts in the application area of smart manufacturing. In this article, we will introduce and explain our 2-step approach. Thereby we will also incorporate some of the results from Yang [11] as he has shown some drawbacks of Kalfoglou and Schorlemmer's approaches and we will use his simpler approach, which enables us to construct the IF channel based on the initial correspondences directly.

## II. APPROACH AND HYPOTHESES

One of the main theories behind the worldwide ‘Factory of the Future’ movement (also known as Smart Manufacturing in US or Industrie 4.0 in Germany) is the idea that machines and systems are getting far more interconnected with each other as they are at the moment to finally reach a level of self-autonomy. One example of this intended form of self-autonomy is the situation where the production system automatically assigns specific manufacturing equipment to the necessary next production step. Within this scenario, the overall system has to be able to decide which manufacturing equipment can conduct a specific production step. New in this scenario is that due to the market pressure and the increasing personalization of products it is expected that the lot sizes in many manufacturing areas will decline. As a result, new products with new specifications and thus new production steps have to be manufactured quite often. A situation where the production system has to decide dynamically which manufacturing equipment will be able to conduct so far not known production steps adequately. We propose that the decision on which machine fits best to a specific production step has to be met on the semantic level. Because this decision depends heavily on the matching of the properties that describe on the one side the production step, like dimensions, material, surface requirements, or delivery date for the product. On the other side, the properties describe the machine capabilities, like vertical or horizontal travelling distances, tools, or capacities of an individual machine. These properties are typically stored in different data sources with different data description languages and schemas. A list of production steps may be saved for example in a relational database, and a description of the available machinery and their capabilities may be stored with the help of a semantic web ontology. This situation describes a highly dynamic environment. The production system is not only faced with so far not known products and thus new production steps, but also with new equipment that might be brought to the shop floor or might be used in cloud-based manufacturing platforms that support new forms of collaboration within the supply chain, e.g. [12] and [13].

This scenario with a list of product specifications stored in one system and a list of production capabilities in another system can perfectly be seen as a kind of distributed system as they are described in the theory of information flow [1] and in the applications of this theory for example in Kalfoglou’s and Schorlemmer’s work as it is shown for example in [4], [6], or in [14].

A core component in the application of the theory of Information Flow based on the work of Kalfoglou and Schorlemmer is the representation of a component’s context in the distributed system with the help of an IF classification. This IF classification is a simple structure that is normally described with the help of a simple table where tokens are listed in rows and types are shown in columns. The examples in the literature so far mostly show quite simple context representations with only a

few types that help to classify the tokens and to give simple insights in a few aspects of the properties of the tokens and the relations between the types and the tokens in question. Unfortunately, this simple table structure is not appropriate for our application domain manufacturing and the quite complex examples that arise from this area. It does not seem feasible for example to put all necessary descriptions about the required properties for all imaginable products and production steps into one single IF classification to prepare a proper decision for a matchmaking with a machine that might be able to conduct a specific production step. What seems to be needed is an alternative method that helps to describe all the necessary properties of the objects (tokens) within the contexts of the products and the production machinery for our application example. To overcome this shortcoming, we propose our 2-step approach, and we are convinced that this approach will help to reduce the complexity that arises when all the context description is given in a single table. Furthermore, we will demonstrate that our approach is able to show that the theory of Information Flow can also be used in the application of more complex examples than those shown in the literature so far.

The novel approach here is the introduction of our 2-step approach on semantically aligning two contexts in a distributed system. According to Kalfoglou and Schorlemmer’s approaches for applying the theory of Information Flow, the types and tokens of an IF classification are represented and described with the help of simple tables – see Table 1 for a typical example taken from [6].

Table 1. Typical representation of IF classification as a simple table

	AG	PA	IND	FS	EUBD
$r_1$	1	1	0	0	0
$r_2$	1	0	1	0	0
$r_3$	1	0	0	0	0
$r_4$	0	0	0	1	1
$r_5$	0	0	0	1	0

In this example, the five tokens ‘ $r_1$  to  $r_5$ ’ are responsibilities that are classified to ministry units (the types), like PA for Passport Agency or IND for Immigration and Nationality Directorate. It can be said that within this example, we are dealing with a well-defined set of ministry units that are used to clarify the relations between responsibilities and the ministry units.

Unfortunately, the situation in our application domain of manufacturing is not that easy. In our production domain, we might be faced with a huge number of fast-changing products and each product might be produced in a sequence of several different production steps. Furthermore, each single production step (our tokens) might be described with some properties (our types) that might be completely different from other production steps. Such properties of a specific production step might be

dimensions and tolerances, materials, surface requirements, necessary production techniques, and maybe a series of other properties that are very specific for each single production step. If we used a table like the one from Table 1 to classify all production steps (tokens) with all necessary properties (types) for each of the production steps and all products, we would end up in a very complex table with a huge number of very different types. The same would be true for the IF classification of the machinery that describes the capabilities of the available production equipment and their tools.

We assume that putting all necessary information for all the production steps or all the production equipment into one single IF classification is very complicated and maybe not even feasible. However, it is of great importance for our work to find a solution for the comprehensive description of production steps and production equipment in IF classifications. IF classifications are a core component for the construction of IF channels, which leads us to the constraints that we need for establishing semantic interoperability.

### III. THE 2-STEP APPROACH

We propose a 2-step approach for the overall matching between production steps and production equipment to reduce the complexity in building up IF classifications, which are basic components of IF. We will do this by splitting the modelling of the overall specifications for a specific production step and the capabilities of the production equipment in several classifications. Thereby the token information will be split over several classifications, which will be processed one after another. While introducing a 2-step matching approach, the first step within this 2-step approach acts like a filter mechanism that selects only the relevant production processes and consequently only implicitly the machinery that is related to the selected production processes. Our 2-step approach will in the first step match between production steps and production processes. In the second step, we will match between detailed descriptions of the specific production steps and the selected production processes (the results of step #1) together with the capabilities of the production equipment that is related to the resulting production processes from step 1.

While matching between production steps and production processes in step #1, only a high-level description in varying details of the production steps and the production processes is given. The result of this matching process is a relation between a specific production step and one or more suitable production processes.

In step #2 the final decision for the best fitting production equipment is met by matching between detailed requirements for the specific production step together with the corresponding production process (results from step #1) and the machines that belong to the resulting production processes from step #1. This matching can be done on more specific details as we are now dealing only with a subset of the machines and tools

from the available machinery based on the results of the preselection from step #1. We propose an approach where each of the detailed descriptions of the production steps and the machine categories is encapsulated in their own IF classifications. There might be for example an IF classification for the detailed description of a production step “making hole” and a separate one for the detailed description of a production step “trimming workpiece” or e.g. an IF classification for drilling machines and a separate one for assembly robots or 3D printers, see Figure 1. These separate IF classifications result in very specific types describing for example only the typical capabilities of a specific machine category. This way, in step #2 also the production step can be described in more detail than in step #1.

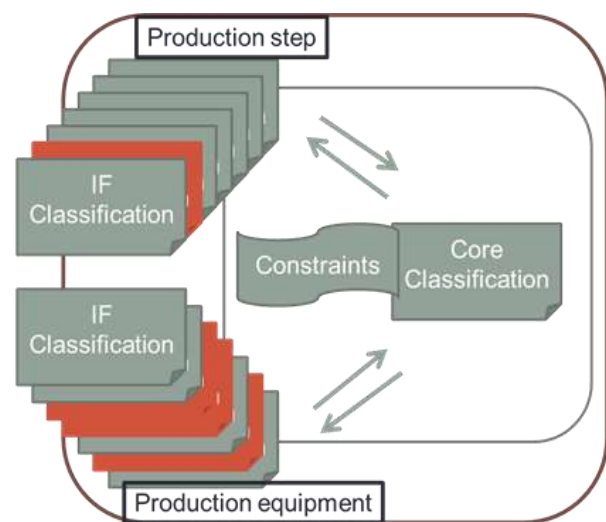


Fig. 1. Matching on subsets in step #2 after the preselection of relevant production processes in step #1

The described 2-step approach is based on the semantic integration with the help of the construction of two separate channels. These two channels are different but related as the participating IF classifications in the channels are related. Within our research work, we will name such related environments composite channels.

### IV. FIRST STEP: THE SELECTION OF SUITABLE PRODUCTION STEPS

In the first step, we are aiming to find out which production processes are suitable for specific production steps. We do this by matching the production steps with production processes to achieve the suitable pairings between them. Therefore, both of these contexts are represented in IF classifications in which the instances of the production steps and the production processes, the so-called tokens, are described by some high-level characteristics, the so-called types. This way, we get an IF classification *PS* for the context representation of the production steps and an IF classification *PP* for the context representation of the production processes.

Usually, a product is produced in a sequence of several production steps. We can address this for example by

indexing the tokens of the IF classification as  $PS_{1-1}$ ,  $PS_{1-2}$ ,  $PS_{1-3}$ ,  $PS_{1-4}$  whereby the first index number represents the number of the product in question and the second one is for a specific production step for the overall product. We will take this indexing approach to create our model of a product that exists of a sequence of necessary production steps.

A production process describes a procedure that is used to create a specific, well-defined geometrical shape and specified characteristics of a workpiece – several production steps and their corresponding production processes lead at the end to the overall product. DIN 8580 [15] lists a taxonomy of major groups of production

processes, and further DIN writings are mentioned in which the subgroups of these major groups are described. We will use parts of these production processes and their subgroups as well as the descriptions and definitions of these production processes in the following examples.

To start with, we are describing a context for the products and their production steps  $PS$  based on our high-level model. The following table shows an example of an IF classification representing some products modelled with the help of their necessary production steps as well as some high-level descriptions (types) for those production steps:

Table 2. Context for products and their production steps – IF classification  $PS$

Production Step $PS$	Trimming workpiece	Deburring workpiece	Making hole	Deburring hole	Finishing workpiece	Assembling	Quality control
$PS_{1-1}$	1	0	0	0	0	0	0
$PS_{1-2}$	0	1	0	0	0	0	0
$PS_{1-3}$	0	0	1	0	0	0	0
$PS_{1-4}$	0	0	1	0	0	0	0
$PS_{1-5}$	0	0	0	1	0	0	0
$PS_{1-6}$	0	0	0	1	0	0	0
$PS_{2-1}$	1	0	0	0	0	0	0
$PS_{2-2}$	0	1	0	0	0	0	0
$PS_{2-3}$	0	0	1	0	0	0	0
$PS_{2-4}$	0	0	0	1	0	0	0

What we see in Table 2 is the description of two products in which the first product has to be conducted with a sequence of six production steps and the second product has a sequence of four production steps. These are the tokens of the classification. The types of the classifications describe which kind of activity has to be executed in the specific production steps.

Now we will describe the context of the production processes, in which we are describing only a selection of production processes in the following table that shows the IF classification  $PP$ :

Table 3. Context for production processes – IF classification  $PP$

Production Process $PP$	Drilling	Gluing	Grinding <sub>co</sub>	Milling	Nailing	Paintin <sub>co</sub>	Reamin <sub>co</sub>	Rivetin <sub>co</sub>	Sanding	Sawing	Screwwin <sub>co</sub>	Sinking	Turning	Weldin <sub>co</sub>
PP1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
PP2	0	1	0	0	0	0	0	0	0	0	0	0	0	0
PP3	0	0	1	0	0	0	0	0	0	0	0	0	0	0
PP4	0	0	0	1	0	0	0	0	0	0	0	0	0	0
PP5	0	0	0	0	1	0	0	0	0	0	0	0	0	0
PP6	0	0	0	0	0	1	0	0	0	0	0	0	0	0
PP7	0	0	0	0	0	0	1	0	0	0	0	0	0	0
PP8	0	0	0	0	0	0	0	1	0	0	0	0	0	0
PP9	0	0	0	0	0	0	0	0	1	0	0	0	0	0
PP10	0	0	0	0	0	0	0	0	0	1	0	0	0	0
PP11	0	0	0	0	0	0	0	0	0	0	1	0	0	0
PP12	0	0	0	0	0	0	0	0	0	0	0	1	0	0
PP13	0	0	0	0	0	0	0	0	0	0	0	0	1	0
PP14	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Table 3 depicts a series of fourteen production processes, whereby the types are describing the specific kind of production process for each of the fourteen tokens. For this example, we assume that the production processes from classification *PP* represent all those processes that can be performed in a specific environment based on the available equipment.

Now that we have represented the IF classifications of both components of our distributed system, namely the context for the production steps *PS* and the context for the production processes *PP*, we will semantically align them to find out, what production processes suit to the specific production steps. Therefore we need a starting point, the initial correspondences.

In Table 2, we have described our context of products with a model of two different products, each of them having a sequence of several production steps. To demonstrate, how this semantic alignment works, we are now focusing on the first product and its production steps  $PS_{1-x}$ . Furthermore, we are having a look on the so far known correspondences between the necessary production steps from *PS* and the available production processes from *PP*, the so-called initial correspondences.

Initial correspondences are the result of a priori knowledge or other kind of heuristics, and for this example, we get our a priori knowledge from some successfully produced products P1 with its production steps  $PS_{1-1} \dots PS_{1-6}$  that had been produced in different ways with partly different production processes. It does not matter for this example, where exactly the initial correspondences come from, but what is really important, is to understand what role correspondences play and how important they are for the resulting semantic alignment.

In IF theory, a correspondence is a pair of tokens from the corresponding IF classifications that describes a particular relationship between the two component classifications. In our case, this means for step #1 that a specific production process from *PP* is suitable to conduct the specific production step from *PS* by taking into account the corresponding types that are used to describe the tokens in question. This pair of tokens can then be used to build up a token  $t_x = \langle t_{i,PS}; t_{j,PP} \rangle$  of the core that is described by the types of the involved tokens from the IF classifications *PS* and *PP* that build up this token  $t_x$ . Given this, it can be derived that an initial correspondence is describing a particular under the conditions of the types of the corresponding component classifications.

In our example, we have the following initial correspondences on token-level:

- (a)  $PS_{1-1} \leftrightarrow PP10$
- (b)  $PS_{1-2} \leftrightarrow PP3$
- (c)  $PS_{1-3} \leftrightarrow PP1$
- (d)  $PS_{1-3} \leftrightarrow PP4$
- (e)  $PS_{1-4} \leftrightarrow PP1$
- (f)  $PS_{1-4} \leftrightarrow PP4$
- (g)  $PS_{1-5} \leftrightarrow PP12$
- (h)  $PS_{1-6} \leftrightarrow PP12$

We can derive from these initial correspondences that for this product P1 the production steps  $PS_{1-3}$  and  $PS_{1-4}$  (both production steps are described by type "Making hole") can be conducted either by production process PP1 (drilling) or by production process PP4 (milling). This is a crucial perception from the manufacturing area that some production steps can be executed via different production processes.

Based on these initial correspondences, we can now construct the IF channel for our product P1 based on the IF classifications and the corresponding infomorphisms. How this process is done is well described for example in the work of Kalfoglou and Schorlemmer, e.g. in [4] or [6] as well as in the literature from other researchers, who mainly rely on the results of Kalfoglou and Schorlemmer's work on IF, e.g. [8], [9], or [10]. While constructing our channel, we will follow the way that is described in chapter 6 of Yang's thesis [11], as Yang found out a way to construct the IF channel out of initial correspondences and IF classifications that is in fact based on the IF application approach of Kalfoglou and Schorlemmer, but not so complex and thus not so error-prone. According to Yang [11, p.94], the way to achieve semantic alignment between two components of a distributed system based on initial correspondences on token-level follows the following three steps:

- (1) Each component has to be captured by an IF classification – for us, these IF classifications are *PS* and *PP*;
- (2) Constructing a core for an IF channel where the tokens are the pairs resulting from the initial correspondences between our components *PS* and *PP* and types are the types in component classifications that are associated with the tokens;
- (3) Completing the IF channel, and identifying all the constraints whereby to identify alignment.

As we already have the classifications that we need for step (1), namely *PS* and *PP*, we can now go over to step (2). As we said previously, the correspondences are used to build the tokens of the core of the IF channel and thereby all the types of the corresponding tokens from the component classifications are building up the types of the core. With our initial correspondences, we come up with the constructed core *PS-PP* for our IF channel that is depicted in Table 4.

What we can now derive within step (3) from this complete channel for product P1, is the constraints that help us to identify the semantic alignment between production steps and production processes and thus the rules with which we are able to find further semantic alignments for other products and their production steps. Firstly we are showing the derived constraints from core *PS-PP* by having a look on the tokens and the relationships between the types describing those tokens:

Trimming workpiece	}	Sawing;
Deburring workpiece	}	Grinding;
Making hole	}	Drilling;

Making hole | Milling;  
 Making hole | Drilling, Milling;

Deburring hole | Sinking;

Table 4. Constructed core *PS-PP* of the IF channel for the product P1

PS-PP	types from classification PS				types from classification PP				
	Trimming workpiece	Deburring workpiece	Making hole	Deburring hole	Drilling	Grinding	Milling	Sawing	Sinking
<PS <sub>1-1</sub> ;PP10>	1	0	0	0	0	0	0	1	0
<PS <sub>1-2</sub> ;PP3>	0	1	0	0	0	1	0	0	0
<PS <sub>1-3</sub> ;PP1>	0	0	1	0	1	0	0	0	0
<PS <sub>1-3</sub> ;PP4>	0	0	1	0	0	0	1	0	0
<PS <sub>1-4</sub> ;PP1>	0	0	1	0	1	0	0	0	0
<PS <sub>1-4</sub> ;PP4>	0	0	1	0	0	0	1	0	0
<PS <sub>1-5</sub> ;PP12>	0	0	0	1	0	0	0	0	1
<PS <sub>1-6</sub> ;PP12>	0	0	0	1	0	0	0	0	1

This is what we have learned from the successful production of product P1 with partly different production processes:

The successful rules for the production of product P1 say that to trim a workpiece on length (PS1-1) we can do it by sawing; the deburring (PS1-2) can be done by grinding; the holes (PS1-3 and PS1-4) can be put into the workpiece by either drilling or milling; and the deburring of the holes (PS1-5 and PS1-6) can be done by sinking.

We can now adapt these rules to the second product P2 and its production steps PS2-x from our classification PS. By applying these rules based on the constraints from core PS-PP from Table 4, we get the following initial correspondences for the production steps PS2-x from product P2:

- Trimming workpiece (PS<sub>2-1</sub>) <-> Sawing (PP10)
- Deburring workpiece (PS<sub>2-2</sub>) <-> Grinding (PP3)
- Making hole (PS<sub>2-3</sub>) <-> Drilling (PP1)
- Making hole (PS<sub>2-3</sub>) <-> Milling (PP4)
- Deburring hole (PS<sub>2-4</sub>) <-> Sinking (PP12)

This brings us to the new core *PS-PP*<sub>2</sub> showing all the core tokens for the production processes PS<sub>1-x</sub> and PS<sub>2-x</sub> from product P1 and P2:

What we have reached now are the constraints taken out of core PS-PP which give us a set of rules that help us to easily filter out which kind of production processes might be suitable for known or so far unknown production steps PS<sub>x-y</sub> from known or so far unknown products P<sub>x</sub>. These rules are valid as a filtering mechanism until we learn that a specific production step should not be done by a specific production process (i.e. a specific kind of hole should not be produced by milling) or a specific production step can also be done by another or maybe a new production process (i.e. trimming a workpiece by either sawing or shearing).

Now that we know what kind of production processes might be suitable for specific production steps, we want to find out, which kind of equipment that is related to a derived production process is suitable for a specific production step. This brings us to step #2 of our 2-step approach for semantically aligning product specifications with manufacturing equipment capabilities.

Table 5. Constructed core  $PS-PP_2$  of the IF channel for the product P1 and P2 and their production steps

PS-PP <sub>2</sub>	types from classification PS										types from classification PP						
	mi ng	wo rkp	im ep	wo rkp	Ma kin	ep hol	blu r	im ep	hol	Dr llin	ep	Gr ndi ng	Mi llin	ep	Sa wi ng	Si n kin	ep
<PS1-1;PP10>	1		0		0			0		0		0		1		0	
<PS1-2;PP3>	0		1		0			0		0		1		0		0	
<PS1-3;PP1>	0		0		1			0		1		0		0		0	
<PS1-3;PP4>	0		0		1			0		0		0		1		0	
<PS1-4;PP1>	0		0		1			0		1		0		0		0	
<PS1-4;PP4>	0		0		1			0		0		0		1		0	
<PS1-5;PP12>	0		0		0			1		0		0		0		0	
<PS1-6;PP12>	0		0		0			1		0		0		0		0	
<PS2-1;PP10>	1		0		0			0		0		0		1		0	
<PS2-2;PP3>	0		1		0			0		0		1		0		0	
<PS2-3;PP1>	0		0		1			0		1		0		0		0	
<PS2-3;PP4>	0		0		1			0		0		0		1		0	
<PS2-4;PP12>	0		0		0			1		0		0		0		0	

V. SECOND STEP: THE SELECTION OF SUITABLE EQUIPMENT

Now, in the second step, we want to find out, which equipment might be used to conduct a specific production step. Other than in step #1, we are now matching the production steps with the available equipment and tooling. Thereby we are using a more detailed description of the production steps than in step #1 showing more specific requirements for this specific production step of the overall product. We are also using a specific description for the equipment that is related to the selected production processes for this production step from step #1.

One result of our example from step #1 was to get the following constraints for the production steps “Making hole”:

- Making hole | Drilling;
- Making hole | Milling;
- Making hole | Drilling, Milling;

From these constraints, we can see that a hole could be done either by drilling or milling. We will take this outcome as a baseline for the construction of the channel in step #2 to derive the semantic alignment between specific production steps and the equipment and tooling that is related to the selected production processes from step #1.

The construction of the IF channel is done quite similarly to the construction of the IF channel in step #1. Firstly, we are describing the context for both components of our distributed system, namely the context for the detailed description of the production steps *DD* and the context for the equipment and tooling *ET* that is related to the selected production processes from step #1. Then, we will construct the core of the channel and finally derive the constraints from the resulting IF channel.

We start with the context of the detailed description of the production step *DD*, in which we are focusing on the production steps  $PS_{1-3}$ ,  $PS_{1-4}$ , and  $PS_{2-3}$ , which are the production steps for “making hole”:

Table 6. Classification *DD* for the detailed description of the production steps “making hole”

DD <sub>MakingHole</sub>	Requirements/Specifications, linear dimensions all in millimetre (mm)				
	Metal	Wood	Long hole	Hole diameter ≤ 13	Hole diameter > 13
Production Step PS <sub>1-3</sub>	1	0	0	1	0
PS <sub>1-4</sub>	1	0	0	0	1
PS <sub>2-3</sub>	1	0	1	1	0

In Table 6 the production steps for the type “making hole” are described in more detail by giving information on the material that has to be used or the hole is a long hole and whether the diameter of the hole is greater or smaller than 13mm. For this example, we do not need the exact coordinates or the exact diameters of the holes as we are not aiming to go down to the process level. For the

demonstration of our 2-step approach, this is a handy and sufficient description of the production step.

The second component of our distributed system is the context of the equipment and tooling for the selected production processes drilling and milling that had been filtered out in step #1:

Table 7. Classification  $ET_{Drilling}$  for the context production process drilling

$ET_{Drilling}$	Capabilities, linear dimensions all in millimetre (mm)				
	Equipment / Tooling - ET	HSS toolset	Wood drill set	Drill chuck $\leq$ 13	Drill chuck $>$ 13
$ET_{D1-1}$		1	0	1	0
$ET_{D1-2}$		1	0	0	1
$ET_{D2-1}$		1	0	1	0
$ET_{D3-1}$		1	0	1	0
$ET_{D3-2}$		1	0	0	1
$ET_{D4-1}$		0	1	1	0
$ET_{D4-2}$		1	0	1	0
$ET_{D5-1}$		1	0	1	0

Table 7 describes the available drilling machines  $ET_{D1} \dots ET_{D5}$  giving information for example on whether drills for specific materials are available or whether there are drill chucks that are able to hold small or big drills. A machine can be equipped with one tooling on a specific time and with some other tooling on another time. This is taken into account by indexes like  $ET_{D1-1}$  and  $ET_{D1-2}$  that indicate for example that the setup for  $ET_{D1-1}$  includes a drill chuck for drills with a diameter up to 13mm and that

the setup for  $ET_{D1-2}$  includes a drill chuck for drills with a diameter greater than 13mm.

For the milling equipment, we have a similar context with slightly different types describing the capabilities of the milling equipment to produce holes. Similar to the drilling machines the milling machines can be equipped with different tooling to address different tasks; this is also indicated by indexes:

Table 8. Classification  $ET_{Milling}$  for the context production process milling

$ET_{Milling}$	Capabilities, linear dimensions all in millimetre (mm)							
	Equipment / Tooling - ET	HSS toolset	Wood drill set	Drill chuck $\leq$ 13	Drill chuck $>$ 13	End mill cutter set	Tool holder $\leq$ 13	Tool holder $>$ 13
$ET_{M1-1}$		1	0	1	0	0	0	0
$ET_{M1-2}$		1	0	0	1	0	0	0
$ET_{M1-3}$		0	1	1	0	0	0	0
$ET_{M1-4}$		0	1	0	1	0	0	0
$ET_{M1-5}$		0	0	0	0	1	1	0
$ET_{M1-6}$		0	0	0	0	1	0	1
$ET_{M2-1}$		1	0	1	0	0	0	0
$ET_{M2-2}$		0	0	0	0	1	1	0
$ET_{M2-3}$		0	0	0	0	1	0	1
$ET_{M3-1}$		0	0	0	0	1	1	0
$ET_{M3-2}$		0	0	0	0	1	0	1



With these separate classifications for the equipment and tooling contexts, we are able to give more specific information on a specific group of machinery that is related to the selected production processes from step #1. If we had not this 2-step approach, we would have had to incorporate all the specific types for all possible equipment groups based on the available production processes into one single table in the classification *PP*. This would be way too many types and makes the whole process for the construction of the IF channel complex and error-prone. With our 2-step approach, we can avoid this complexity and, what is more important, we can build up a detailed description for the contexts of production steps *DD* and the related contexts of equipment and tooling *ET<sub>x</sub>*.

Now we have to construct the IF channel out of the IF classifications and given initial correspondences to derive the constraints that help us to reach the semantic alignment between product specifications and equipment

capabilities. We will assume again that we have the initial correspondences from our a priori knowledge based on the successful production of product P1 and P2 including the successful production of the production steps *PS<sub>1-3</sub>*, *PS<sub>1-4</sub>*, and *PS<sub>2-3</sub>* (making hole). For our example, we assume that we have learned about the following initial correspondences so far:

- (a)  $PS_{1-3} \leftrightarrow ET_{D1-1}$
- (b)  $PS_{1-3} \leftrightarrow ET_{D4-2}$
- (c)  $PS_{1-3} \leftrightarrow ET_{M1-1}$
- (d)  $PS_{1-4} \leftrightarrow ET_{D1-2}$
- (e)  $PS_{1-4} \leftrightarrow ET_{M1-2}$
- (f)  $PS_{2-3} \leftrightarrow ET_{M1-5}$

Based on our classifications *DD*, *ET<sub>Drilling</sub>*, and *ET<sub>Milling</sub>* as well as our initial correspondences, we can construct the following core of our IF channel for the drilling and milling machinery:

Table 9. Constructed core *DD-ET<sub>x</sub>* for the contexts making hole and drilling/milling

types from classification DD	types from classification														
	<i>ET<sub>Drilling</sub></i>				<i>ET<sub>Milling</sub></i>										
<i>DD-ET<sub>x</sub></i>	Metal	Long hole	Hole diameter ≤ 13	Hole diameter > 13	HSS toolset	Wood drill set	Drill chuck ≤ 13	Drill chuck > 13	HSS toolset	Wood drill set	Drill chuck ≤ 13	Drill chuck > 13	End mill cutter set	Tool holder ≤ 13	Tool holder > 13
<PS <sub>1-3</sub> ;ET <sub>D1-1</sub> >	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0
<PS <sub>1-3</sub> ;ET <sub>D4-2</sub> >	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0
<PS <sub>1-3</sub> ;ET <sub>M1-1</sub> >	1	0	1	0	0	0	0	0	1	0	1	0	0	0	0
<PS <sub>1-4</sub> ;ET <sub>D1-2</sub> >	1	0	0	1	1	0	0	1	0	0	0	0	0	0	0
<PS <sub>1-4</sub> ;ET <sub>M1-2</sub> >	1	0	0	1	0	0	0	0	1	0	0	1	0	0	0
<PS <sub>2-3</sub> ;ET <sub>M1-5</sub> >	1	1	1	0	0	0	0	0	0	0	0	0	1	1	0

From this core, we can derive for example the following constraints based on our initial correspondences:

- Metal, Hole diameter ≤ 13 ⊢ Drilling{HSS toolset, Drill chuck ≤ 13};
- Metal, Hole diameter ≤ 13 ⊢ Milling{HSS toolset, Drill chuck ≤ 13};
- Metal, Hole diameter > 13 ⊢ Drilling{HSS toolset, Drill chuck > 13};
- Metal, Hole diameter > 13 ⊢ Milling{HSS toolset, Drill chuck > 13};
- Metal, Hole diameter ≤ 13 ⊢ Drilling{HSS toolset, Drill chuck ≤ 13}, Milling{HSS toolset, Drill chuck ≤ 13};
- Metal, Hole diameter > 13 ⊢ Drilling{HSS toolset, Drill chuck > 13}, Milling{HSS toolset, Drill chuck > 13};

Metal, Long hole, ≤ 13 ⊢ Milling{End mill cutter set, Tool holder ≤ 13};

- Long hole ⊢ Milling;
- Long hole, Drilling ⊢;
- Metal ⊢ HSS tool set;
- Long hole ⊢ End mill cutter set;
- Hole diameter ≤ 13 ⊢ Drill chuck ≤ 13;
- Hole diameter > 13 ⊢ Drill chuck > 13;

We can now use these constraints for a so far unknown product P3 and its production steps *PS3-3*, *PS3-4*, and *PS3-5*, which are all standing for making holes. We can use these constraints to crosscheck our 2-step approach for semantically aligning product specifications with production capabilities by matching specific production steps with available equipment and tooling.

Table 10 shows the production steps for making hole from product P1, P2, and the new product P3:

Table 10. Classification *DD* for making hole for products P1, P2, and P3

Production Step PS	Requirements/Specifications, linear dimensions all in millimetre (mm)				
	Metal	Wood	Long hole	Hole diameter $\leq$ 13	Hole diameter $>$ 13
PS <sub>1-3</sub>	1	0	0	1	0
PS <sub>1-4</sub>	1	0	0	0	1
PS <sub>2-3</sub>	1	0	1	1	0
PS <sub>3-3</sub>	1	0	0	1	0
PS <sub>3-4</sub>	1	0	0	0	1
PS <sub>3-5</sub>	1	0	1	1	0

Table 11. Constructed core *DD-ET<sub>x</sub>* for product P3 and its production processes

DD-ET <sub>x</sub>	types from classification DD				types from classification ET <sub>Drilling</sub>				types from classification ET <sub>Milling</sub>							
	Metal	Long hole	Hole diameter $\leq$ 13	Hole diameter $>$ 13	HSS toolset	Wood drill set	Drill chuck $\leq$ 13	Drill chuck $>$ 13	HSS toolset	Wood drill set	Drill chuck $\leq$ 13	Drill chuck $>$ 13	End mill cutter set	Tool holder $\leq$ 13	Tool holder $>$ 13	
<PS <sub>3-3</sub> ;ET <sub>D1-1</sub> >	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	
<PS <sub>3-3</sub> ;ET <sub>D2-1</sub> >	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	
<PS <sub>3-3</sub> ;ET <sub>D3-1</sub> >	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	
<PS <sub>3-3</sub> ;ET <sub>D4-2</sub> >	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	
<PS <sub>3-3</sub> ;ET <sub>D5-1</sub> >	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	
<PS <sub>3-3</sub> ;ET <sub>M1-1</sub> >	1	0	1	0	0	0	0	0	1	0	1	0	0	0	0	
<PS <sub>3-3</sub> ;ET <sub>M2-1</sub> >	1	0	1	0	0	0	0	0	1	0	1	0	0	0	0	
<PS <sub>3-4</sub> ;ET <sub>D1-2</sub> >	1	0	0	1	1	1	1	1	0	0	0	0	0	0	0	
<PS <sub>3-4</sub> ;ET <sub>D3-2</sub> >	1	0	0	1	1	1	1	1	0	0	0	0	0	0	0	
<PS <sub>3-4</sub> ;ET <sub>M1-2</sub> >	1	0	0	1	0	0	0	0	1	0	0	1	0	0	0	
<PS <sub>3-5</sub> ;ET <sub>M1-5</sub> >	1	1	1	0	0	0	0	0	0	0	0	0	1	1	0	
<PS <sub>3-5</sub> ;ET <sub>M2-2</sub> >	1	1	1	0	0	0	0	0	0	0	0	0	1	1	0	
<PS <sub>3-5</sub> ;ET <sub>M3-1</sub> >	1	1	1	0	0	0	0	0	0	0	0	0	1	1	0	

The successful rules for making holes for product P1 and product P2 say that to drill a hole in a metal workpiece with a diameter  $\leq$  13mm (PS<sub>3-3</sub>), we need either a drilling machine with an HSS tool set and a drill chuck  $\leq$  13mm or a milling machine with an HSS tool set and a drill chuck  $\leq$  13mm. Furthermore, we have learned from the constraints that to make a hole in a metal workpiece with a diameter  $>$  13mm (PS<sub>3-4</sub>), we need either a drilling machine with an HSS tool set and a drill chuck  $>$  13mm or a milling machine with an HSS tool set and a drill chuck  $>$  13mm. To produce a long hole in a metal workpiece with a diameter  $\leq$  13mm (PS<sub>3-5</sub>), we can derive from the constraints that we need a milling machine with an end mill cutter set and a tool holder  $\leq$  13mm.

This brings us to the following initial correspondences for product P3 and its production steps PS<sub>3-3</sub>, PS<sub>3-4</sub>, and PS<sub>3-5</sub> based on the learnings from the successful

production of product P1 and P2 and the classifications ET<sub>Drilling</sub> and ET<sub>Milling</sub> from Table 7 and Table 8, respectively:

- (a) PS<sub>3-3</sub> <-> ET<sub>D1-1</sub>
- (b) PS<sub>3-3</sub> <-> ET<sub>D2-1</sub>
- (c) PS<sub>3-3</sub> <-> ET<sub>D3-1</sub>
- (d) PS<sub>3-3</sub> <-> ET<sub>D4-2</sub>
- (e) PS<sub>3-3</sub> <-> ET<sub>D5-1</sub>
- (f) PS<sub>3-3</sub> <-> ET<sub>M1-1</sub>
- (g) PS<sub>3-3</sub> <-> ET<sub>M2-1</sub>
- (h) PS<sub>3-4</sub> <-> ET<sub>D1-2</sub>
- (i) PS<sub>3-4</sub> <-> ET<sub>D3-2</sub>
- (j) PS<sub>3-4</sub> <-> ET<sub>M1-2</sub>
- (k) PS<sub>3-5</sub> <-> ET<sub>M1-5</sub>
- (l) PS<sub>3-5</sub> <-> ET<sub>M2-2</sub>
- (m) PS<sub>3-5</sub> <-> ET<sub>M3-1</sub>

What we can see from these initial correspondences is that we have now far more initial correspondences as a starting point than for product P1 and P2. This is because we have translated our resulting constraints from the core in Table 9 to new initial correspondences for the production steps of product P3 and these initial correspondences might take into account equipment and tooling that might not been used for product P1 and P2, but might be able to do the necessary job.

With these initial correspondences we can now get the following core for product P3:

From this core, we can now derive the same constraints as from the core for product P1 and P2 in Table 9 and these constraints give us the information of which equipment with which kind of tooling is able to conduct a specific production step, which is already known or so far unknown. And this information is what we wanted to receive by semantically aligning the two contexts product specification (production step) and production capabilities (production process and production equipment).

## VI. CONCLUSION

Starting from the observation that the theory of Information Flow (IF) [1] is a mathematical sound theory showing that information can flow between the components of a distributed system, we have recognized that the application approaches of this theory so far only show quite simple examples in the literature, e.g. [6] or [8].

We have shown in this paper that IF can also be applied in more complex environments when the procedure for the semantical alignment based on IF is split into at least two consecutive steps. This approach reduces the complexity while building up the IF classifications that are necessary to build up the IF channels.

To showcase the applicability of our approach in a more complex environment, we have selected an example from the manufacturing area, where product specifications have to be matched to available production capabilities in a specific manufacturing environment.

In the first step of our showcase, we are filtering out which kind of production processes can be used to conduct specific production steps of a product. This is done by building up an IF channel based on some initial correspondences resulting from previous successful production of the products in question and a set of production steps and production processes that are represented in two different IF classifications for the production steps and production processes, respectively.

In the second step, we are matching more detailed descriptions of the production steps in question with the equipment and tooling that is related to the selected production processes from the resulting pairings from step #1. This time, we are able to give more detailed descriptions in the IF classifications of step #2, namely the IF classification for the detailed description of the requirements of the production steps and the IF

classification that represents a detailed description of the available production equipment and tooling.

The result of our showcase based on the composite channels from step #1 and step #2 shows that we can semantically align the components of a distributed system by using IF and the known application approaches from Kalfoglou and Schorlemmer even in a complex environment, when we are splitting the complexity of building up the necessary IF classifications in at least two successive steps with different levels of detail.

## REFERENCES

- [1] J. Barwise and J. Seligman, *Information Flow: The Logic of Distributed Systems*, Reprinted. New York: Cambridge University Press, 1997.
- [2] R. E. Kent, "The Information Flow Foundation for Conceptual Knowledge Organization," in *Sixth International Conference of the International Society for Knowledge Organization*, 2000.
- [3] R. E. Kent, "Semantic Integration in the Information Flow Framework," *Dagstuhl Seminar Proceedings 04391*, pp. 1–12, 2005.
- [4] Y. Kalfoglou and M. Schorlemmer, "IF-Map: An Ontology-Mapping Method based on Information-Flow Theory," *Journal on Data Semantics*, vol. 1, no. 1, pp. 98–127, 2003.
- [5] Y. Kalfoglou and M. Schorlemmer, "Information-Flow-based Ontology Mapping," in *On the Move to Meaningful Internet Systems, 2002 - DOA/CoopIS/ODBASE 2002 Confederated International Conferences DOA, CoopIS and ODBASE 2002*, 2002, pp. 1132–1151.
- [6] M. Schorlemmer and Y. Kalfoglou, "Using Information-Flow Theory to Enable Semantic Interoperability," in *Artificial Intelligence Research and Development*, 2003, pp. 1–6.
- [7] M. Schorlemmer and Y. Kalfoglou, "Progressive Ontology Alignment for Meaning Coordination: an Information-Theoretic Foundation," *Knowledge Creation Diffusion Utilization*, pp. 737–744, 2005.
- [8] K. Xu and J. Feng, "An information flow based approach to finding informational relationships between questionnaires," *IET International Conference on Information Science and Control Engineering 2012 (ICISCE 2012)*, p. 1.04-1.04, 2012.
- [9] G. Yang and J. Feng, "Database Semantic Interoperability based on Information Flow Theory and Formal Concept Analysis," *International Journal of Information Technology and Computer Science*, vol. 4, no. 7, pp. 33–42, Jul. 2012.
- [10] R. Mantri, "Database Web Service Composition based on the Notion of 'Informational Relationship,'" Doctoral Thesis, University of the West of Scotland, 2013.
- [11] G. Yang, "Achieving Database Semantic Interoperability with a Perspective of Information Flow and Formal Concepts," Doctoral Thesis, University of the West of Scotland, 2015.
- [12] X. Xu, "From cloud computing to cloud manufacturing," *Robotics and Computer-Integrated Manufacturing*, vol. 28, no. 1, pp. 75–86, Feb. 2012.
- [13] D. Wu, M. J. Greer, D. W. Rosen, and D. Schaefer, "Cloud manufacturing: Strategic vision and state-of-the-art," *Journal of Manufacturing Systems*, pp. 1–16, May 2013.
- [14] Y. Kalfoglou and M. Schorlemmer, "Formal Support for Representing and Automating Semantic Interoperability,"

in *The Semantic Web: Research and Applications: First European Semantic Web Symposium, ESWS 2004*, 2004, pp. 45–60.

- [15] DIN8580:2003-09, “Manufacturing processes – Terms and definitions, divisions“ (org. “Norm DIN 8580:2003-09, Fertigungsverfahren - Begriffe, Einteilung”), Beuth Verlag GmbH, Berlin.

### Authors’ Profiles



**Andreas Bildstein:** Group Manager of the research team “IT Applications and Services for Production” at the Competence Center Digital Tools for Manufacturing at Fraunhofer IPA, Stuttgart/Germany. Doctorate candidate at the Database and Knowledge Management Research Group in the School of Engineering and Computing at the University of the West of Scotland, Paisley/UK. His research interests include among others Industrie 4.0, Cyber-physical systems for production, cloud-based manufacturing,

Semantic Web Technology for Production, and Semantic Interoperability.



**Junkang Feng BSc, MPhil, PhD:** majoring in Engineering and in Computer Science. He worked as an RA at the University of Manchester UK. Dr Feng is a Senior Lecturer and a member of the Database and Knowledge Management Research Group at the University of the West of Scotland, UK. He is a Visiting Professor at Donghua University and Beijing Union University in China. Dr Feng’s research interests include information and information flow theories, distributed information systems and database theory and technology including NoSQL databases, and has published over 220 research articles. Dr Feng completed/was involved in 9 externally funded research projects. His work was submitted to RAE 2001 and REF 2014 in the UK. He has supervised 8 PhD programmes and 36 MSc dissertation projects to successful completion.

**How to cite this paper:** Andreas Bildstein, Junkang Feng, " A Channel Theory based 2-Step Approach to Semantic Alignment in a Complex Environment", *International Journal of Modern Education and Computer Science(IJMECS)*, Vol.9, No.9, pp.1-12, 2017.DOI: 10.5815/ijmeecs.2017.09.01