Towards Linked Data in Physics

Marcin Skulimowski
Faculty of Physics and Applied Informatics, University of Lodz, Pomorska 149/153, 90-236 Lodz, Poland

Keywords: Semantic Web, Linked Data, Linked Science.

Abstract: Linked Data refers to machine-understandable data published on the Web using the Resource Description Framework (RDF). Publishing and linking data using RDF make data easier to discover and easier to use. Nowadays, Linked Data principles have gained popularity in many fields. In this paper we are going to examine the possibility of obtaining Linked Data in physics. In particular, we discuss the principles of Linked Data in the context of publications on physics and present examples of Linked Data in a branch of physics called quantum mechanics. Moreover we present a web tool supporting manual creation of Linked Data in domains in which automating extraction of the data is not possible.

1 INTRODUCTION

Linked Data refers to machine-understandable data published on the Web using the Resource Description Framework (RDF). The key feature of RDF is that it allows us to link identifiers for entities (e.g. digital objects, real objects, abstract concepts) and not just documents (as HTML). Moreover, RDF links are typed, i.e. the nature of connection between two linked entities can be stated explicitly. Publishing and linking data using RDF will make data on the Web easier to discover, more accessible and, thus, easier to use. Until now, Linked Data principles have gained popularity in many fields and are applied to various kinds of data, e.g. government data (Sheridan and Tennison, 2010), commerce (Hepp, 2008), education and social media (Heath and Bizer, 2011). Applications of Linked Data can be also found in various branches of science, e.g. in geography (Sren Auer and Hellmann, 2009), biomedicine (Ciccarese et al., 2008), chemistry and biology (Wiljes and Cimiano, 2012). The Linked Open Data initiative of the University of Münster (LODUM)[1], which aims to publish any non-sensitive data online according to the Linked Data principles is worth mentioning (Kauppinen et al., 2013).

The purpose of this study is to examine the possibility of obtaining Linked Data in physics. Generally, the data and knowledge are usually contained in research articles and books. They usually involve considerations using advanced mathematical formalism. The data may also be formed into plots and tables (e.g. data from experiments). These ways of storing scientific data implies difficulties in data processing, in particular searching for information. The searching is usually restricted to full text search. Machines are not able to integrate data from the articles. From a machine point of view the articles are data islands. And widely used citations are nothing more than relations between articles. As a consequence a statement that article A cites article B has no precise meaning. Does it mean that in article A the ideas and propositions from article B are creatively developed? Or does it mean that in article A article B is only merely mentioned? To answer these questions we have to look through the article. The question arises then whether it is possible to represent data and knowledge available in physics in such a way that will enable integration, reusing and sharing the data i.e. according to Linked Data principles. In order to answer this question, we focus our attention on a branch of physics called quantum mechanics. The main reason for this choice is central importance of the theory for other branches of physics and technology. Moreover, in order to obtain Linked Data, we need formal representations of domain-specific terms which can be used in RDF descriptions. In the case of quantum mechanics we are developing an ontology which is suitable for this purpose (Skulimowski, 2010).

In the rest of paper, we analyze the principles of Linked Data in the context of data and knowledge available in quantum mechanics. In particular, we give an example of Linked Data for this disci-

[1] http://lodum.de

Skulimowski M..
Towards Linked Data in Physics.
DOI: 10.5220/0004504700480054
In *Proceedings of the 5th International Conference on Computer Supported Education* (CSEDU-2013), pages 48-54
ISBN: 978-989-8565-53-2

pline. We also present a draft version of a web tool shortly called LYR (Link Your Research) which supports manual creation of Linked Data in domains in which automated extraction of the data is not possible (e.g. quantum mechanics). The paper ends with a short discussion and an outline of future work.

2 RELATED WORK

Previous work related to the presented research have already been done. Several approaches for a creation of linked data have been proposed. Among them there are automated approaches and approaches supporting manual creation (Wiljes and Cimiano, 2012). In order to describe data appropriate vocabularies are needed. As a result various ontologies have been developed, e.g. supporting semantic publishing and referencing (SPAR)[2], scientific discourse[3] and open annotations in science[4]. There also exists a number of systems similar (in some sense) to LYR. ScienceWISE Platform allows a community of scientists annotating and bookmarking research articles using appropriate ontologies (Aberer et al., 2011). Annotations can be also created using 4A framework based on idea ”annotations any where, annotations any time” (Smrz and Dytrych, 2011). Scientific workflows and other artefacts can be published, shared and discovered using *my*Experiment Virtual Research Environment (Goble and Roure, 2007).

3 LINKED DATA PRINCIPLES AND QUANTUM MECHANICS

Tim Berners-Lee introduced Linked Data principles describing a set of best practices for publishing structured data on the Web (Berners-Lee, 2006). Let us now consider the principles in the context of data and knowledge contained in research publications on quantum mechanics.

3.1 Naming Things with URI

The first Linked Data principle advocates naming entities (real or abstract) with URIs (Heath and Bizer, 2011). However a question may arise, whether it is possible for entities appearing in research articles on quantum mechanics. In this case the entities usually correspond to some ’elements’ of mathematical structure of the theory e.g. *observable, quantum state, orthogonality*. There are also entities corresponding to the structure of an article e.g. *definition, theorem, lemma*. In order to name these entities with a URI we can utilize the Web accessibility of the research article through the URL. And this URL can be used to obtain URIs for the entities considered in the article. Let us then consider a research article with the following URI:

```
http://example.org/art3
```

We want to assign URIs to various entities from the article. Assume, for example, that some *Concept* is considered in the article. Then it is possible to name this concept with the URI:

```
http://example.org/art3#Concept
```

It happens sometimes that an element considered in an article (e.g. *equation, condition*) has no special name. Inside of the article we can refer to it using its number (e.g. *...equation (12)...*, *...condition (7)...*). This number can be also used to assign a URI to the element e.g.

```
http://example.org/art3#12
```

Standard elements of an article corresponding to its structure (e.g. *theorems, definitions, lemmas* etc.) are usually numbered too. These numbers can be also used in naming the elements with URIs e.g.

```
http://example.org/art3#Theorem_3
http://example.org/art3#Definition_2
```

In addition, elements of the theory considered in articles are very often denoted by symbols e.g. ψ, H_2. In these cases it is possible to assign a URI using (simplified) Latex symbols e.g.

```
http://example.org/art3#Psi
http://example.org/art3#H_2
```

for ψ and H_2.

As the above examples show that it is possible to name various entities from research publications with URIs. Notice, however, that in the case when one entity is named with two different URIs (e.g. using an entity name and an entity number), URI aliases will appear.

3.2 Providing Useful RDF Information

Descriptions of resources that are intended for machines should be represented in RDF. In general, it is an impossible task to represent the whole scientific article on quantum mechanics using RDF triples. This is mainly because of the complicated mathematical

[2]http://purl.org/spar/
[3]http://purl.org/swan/1.2/discourse-relationships/
[4]http://www.purl.org/ao/

formalism used in the theory which obviously cannot be represented in RDF. We may, however, apply a light-weight approach and concentrate only on RDF links between elements of the theory considered in some article and elements of the theory from other articles or publications. Moreover, one may try to determine types of the elements assigning them to vocabulary terms (see Fig.1).

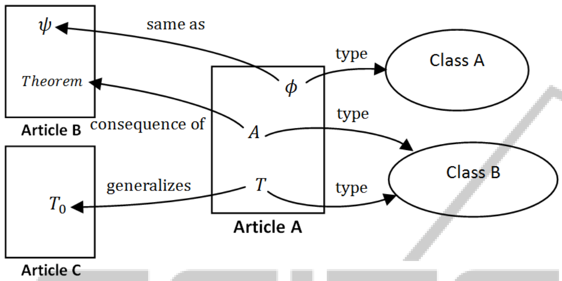

Figure 1: RDF links between entities from article A and entities from articles B and C. The entities from article A are also linked to classes from ontologies.

In order to create RDF links in quantum mechanics we need an appropriate vocabularies. The vocabularies are formally defined in the following ontologies:

- quONTOm[5] - an OWL ontology describing main concepts (e.g. *observable*, *Hamiltonian*, *spin*) and relations (e.g. *commutator*, *orthogonality*) in quantum mechanics. Aside from quantum mechanical concepts and relations, the ontology contains elements which in fact do not belong to the domain of quantum mechanics e.g. mathematical objects which should be contained in an ontology for mathematics. Unfortunately, to our best knowledge, an appropriate ontology for mathematics, which could be imported and used in quONTOm ontology does not exist. However, we think that in the future these concepts will be separated from the ontology and become parts of auxiliary ontology (or ontologies).
- PHYSO (Physical Sciences Ontology)[6] - an ontology describing general concepts (e.g. *principle, problem, assumption*) and relations (e.g. *has property, consequence of*) of physical sciences. The ontology can be used to characterize component parts and relations of any physical theory and not only quantum mechanics.
- SACO (Scientific Article Content Ontology)[7] - an ontology containing a set of objects properties enabling description of what is done/used in a research paper. In such article something (e.g. some element of the theory) is *analyzed*, *described* etc. Equivalently, using 'dummy' subject, we can say that the paper *considers*, *describes* something (Glasman-Deal, 2010).

[5]http://purl.org/quONTOm
[6]http://purl.org/lyr/physo
[7]http://purl.org/lyr/saco

We stress that above ontologies are at the moment incomplete and are gradually developed towards more complete forms. Consequently, parts of them might be changed in the near future.

3.3 Including Links to Other Things

The fourth principle of Linked Data recommends including links to other URIs. These links are crucial because they enable the discovery of additional data resources. There are three types of RDF links.

Relationship Links. In one research article an entity (e.g. *equation*, *concept*, *definition*) from another article can be used (e.g. generalized). Assuming that this element is named with a URI (according to one of the methods considered above) we can create the following RDF link pointing to this element:

```
art3:H phys:generalizes art6:H_0 .
```

A reference to some concept introduced in another article can be represented by the following link:

```
art3:Concept sac:isIntroducedIn
<http://ex2.org/art6> .
```

where `http://ex2.org/ar6` is a URI of the article. Moreover, it is very often that some entity introduced in one article is a solution to a problem considered in another article. We can express this by the following relationship link:

```
art3:10 phys:solutionTo art6:Theorem_5 .
```

Identity Links. An important part of RDF links in Linked Data are identity links. An element of the theory (e.g. *operator*, *concept*, *definition*) may be named with two different symbols in two different articles. In this case we can create the following identity link:

```
art3:V_1 owl:sameAs art6:W_1 .
```

It may happen that the same definition may have two different numbers in different articles. We can represent the situation as follows:

```
art3:Def_1 owl:sameAs art6:Def_2 .
```

Identity links are very important because they enable expression of different views on the same element of the theory named with URI. Moreover, they enable clients to retrieve further descriptions about the element.

Vocabulary Links. The last type of RDF links are vocabulary links pointing to definitions of vocabulary terms. For example, in a research paper on quantum mechanics some observable C_2 can be considered. Then we can create the following link:

```
art3:C_2 rdf:type quo:Observable.
```

It is shown in the next section that RDF links can be used in RDF descriptions of research articles.

4 LINKING RESEARCH DATA

Let us assume now that we want to create RDF links for some research paper on quantum mechanics. This can be done in the following steps:

1. We choose entities from the article which we want to describe in RDF (e.g. *concept 7*, *formula (12)*, Ψ_n etc.).
2. We name the entities with URIs (different ways of creating URIs were presented in the previous section).
3. Next we determine the relation between the paper and the chosen elements (using e.g. SACO ontology).
4. Using appropriate ontologies (e.g. PHYSO, quONTOm), we determine types of chosen elements. In this way we create *vocabulary links*.
5. We add *identity links* pointing at URIs used in other articles to identify the same entities.
6. We create *relationship links* between the entities of the considered article and entities from other articles.

We assume that RDF links obtained initially may be incomplete. The quality and depth of detail can be improved over time by adding new RDF links by the research community. There are many reasons for this 'progressive enrichment' (Dodds and Davis, 2011). For example, the creator of the initial RDF links could not know about some facts. Another reason may be some change or improvement in the vocabulary that was used. It very often happens that results of a paper are used in some other paper published later in the future. Then appropriate RDF links between the two papers should be added.

4.1 Example

Let us now consider an example of Linked Data for the case of quantum mechanics. To this end we take into account a research paper available online at arXiv[8] with the following URL: http://http://arxiv.org/abs/1101.3969v1. The paper is related to so called time operator problem. For convenience, we will use the following prefixes: `sac` - Scientific Article Content Ontology, `quo` - quONTOm Ontology, `dr` - SWAN Discourse relationships vocabulary.

In order to obtain RDF links we follow steps presented above:

1. We want to describe in RDF the following entities from the article: M, $\mathbf{g}_{m,\lambda}$, $\sigma(M)$.
2. URIs assigned to the entities: `<#M>`, `<#g_(m,lambda)>`, `<#sigma(M)>`.
3. Relations between the article and the entities:

```
<> sac:analyzes <#M> .
<> sac:determines <#g_(m,lambda)> .
<> sac:determines <#sigma(M)> .
```

4. Types of the entities:

```
<#M> rdf:type
quo:SelfAdjointOperator .
<#g_(m,lambda)> rdf:type
quo:EigenStates .
<#sigma(M)> rdf:type
quo:OperatorSpectrumBounded .
```

5. Identity link:

```
<#M> owl:sameAs
<http://link.aip.org/link/
doi/10.1063/1.3276419#M_F>
```

6. Relationship links:

```
<#M> quo:hasEigenStates
<#g_(m,lambda)> .
<#M> quo:hasSpectrum <#sigma(M)> .
<#M> sac:introducedIn
<http://link.aip.org/link/
doi/10.1063/1.3276419> .
<#M> dr:relatesTo
<http://arxiv.org/abs/
quant-ph/9611015#Theorem_1> .
```

Further examples of RDF abstracts for papers on quantum mechanics can be found on the *Link Your Research* project website[9].

4.2 Benefits

Let us now present probable benefits of applying Linked Data in quantum mechanic, and in physics

[8] http://arXiv.org

[9] http://www.linkyourresearch.org/

in general. To this end assume that we have some collection of scientific papers on quantum mechanics. Today, only full text searching can be carried out on the papers (eventually, with the help of keywords). In the case when PACS (Physics and Astronomy Classification Scheme)[10] is used, searching for papers by subject is also possible. Assume now that for each paper RDF links are created and stored in a triplestore. Then, using SPARQL endpoint one could make various queries against the dataset. For example, in the case of RDF links created for papers related to time operator problem in quantum mechanics the following queries, among others, could be made:

- *for some entities* e.g. self-adjoint time operators (`quo:SelfAdjointOperator`).
- *for properties of a given entity* e.g. for the type of spectrum (`quo:Spectrum`) of some operator (named with URI).
- *for a relationship between given entities* e.g. for the commutator relation between some two operators (`quo:CommutatorRelation`).
- *for entities with properties similar to a given entity* e.g. for time operators with a given spectrum (`owl:sameAs`).
- *for the article in which some entity was introduced* e.g. some time operator (`quo:introducedIn`).
- *for articles in which some entity is analyzed* e.g. some time operator (`sac:analyzes`).
- *for articles in which some entity is generalized* e.g. some quantum system (`sac:generalizes`).

It is obvious that possibility of asking such queries opens new opportunities for the retrieval of information in quantum mechanics. Scientific papers are no longer only human-readable data islands. They become in some part machine-understandable (at least in some part). There are precise relationships between papers and some entities described in them. Thanks to that as results of queries we will obtain concrete entities from scientific papers and not papers containing some string (as today).

5 LYR WEB TOOL

Linked Data for research papers on quantum mechanics has to be created manually by the research community, in particular by the author of a publication. Automatic creation of valuable and precise relationships between concepts and research papers seems to be out of scope of this domain due to its complexity. However, one can imagine tools supporting and simplifying creation of such data. We have developed a prototype of such web tool called LYR (Link Your Research)[11]. This tool supports creation of RDF links for any kind of publication which is available online. Such a publication is called in LYR a *context*. Each *context* has its URI identifier (usually it is a URL of the publication). A context corresponding to some publication contains RDF links created by the research community. In order to use the tool one has to register and possess a personal URI identifier. After simple registration one can create a new context or add links to already created contexts. The process of adding links has a few steps (including six steps described in Section 4) and consists of filling up a simple form available on the website. To this end one has to know terms from appropriate ontologies (see Fig.2). A user can choose ontologies he wants to use from available list. At present, there are only a few ontologies available however, new ontologies will be successively added (any user can suggest adding new ontology).

All RDF links generated using the LYR tool are stored in Virtuoso Open-Source RDF Triple Store[12]. The dataset obtained so far consists of a few hundred of links (RDF triples) generated for several dozen of research articles. Each article is available on the journal's website or at arXiv. Most of the articles are related to time operator problem in quantum mechanics. The reason for this limitation is the content of the ontology for quantum mechanics. In the actual version of the ontology only the most fundamental concepts and relations of the theory are represented, and the time operator problem is related to such fundamental concepts. The test dataset is gradually enlarged to include more links corresponding to various articles.

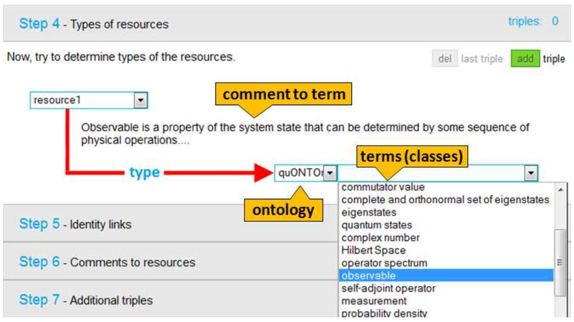

Figure 2: LYR - adding links to a context.

[10] http://publish.aps.org/PACS

[11] The tool will be available soon for tests. Please visit our project website for more information at http://www.linkyourresearch.org

[12] http://virtuoso.openlinksw.com/

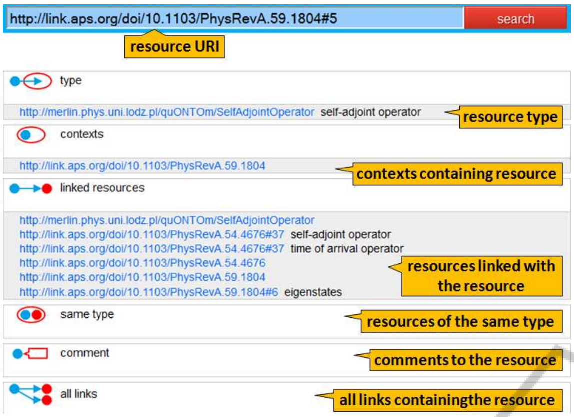

Figure 3: LYR - search for a resource.

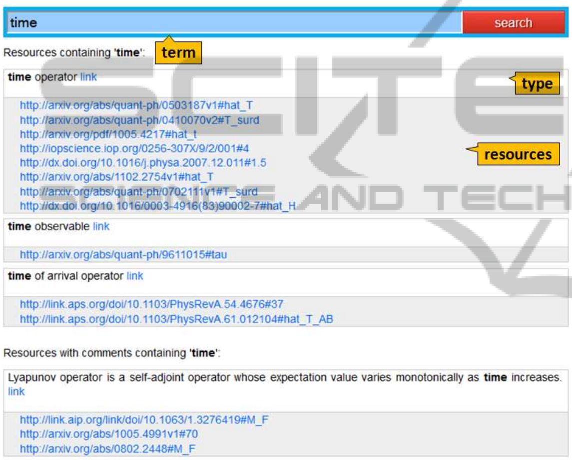

Figure 4: LYR - search for a term.

The LYR tool enables easy search and exploration of the dataset. One can search the dataset for all RDF links from some context, an entity (resource) from some publication (see Fig.3) or a term. In the last case as a result one obtains resources "corresponding" to that term and not strings containing the term (see Fig.4). It is also possible to retrieve links between resources using SPARQL. Finally, it is worth to notice that requesting the linked data corresponding to some context is also possible. To this end one has to use the following URL:

`http://www.linkyourresearch.org/contextURI`

where `contextURI` is the URI of a requested context. When using this URL in a browser a web page containing all RDF triples from the context is loaded. In order to facilitate the creation of RDF links using the LYR tool a Firefox browser extension is developed. The extension will allow to see and add RDF triples for a visited research article (context).

6 DISCUSSION AND FUTURE WORK

In this preliminary paper we consider the Linked Data in the context of physics. In particular, we examine the possibility of representing physics data and knowledge according to the principles. We focus on research articles on quantum mechanics and show that it is possible to obtain Linked Data in those cases, using terms from appropriate ontologies. In particular, it is possible to link elements of the theory from two papers, not just documents containing these papers. There are, however, two important issues deserving attention. First, we want to stress that the obtained RDF links correspond only to part of human-readable data available in the articles. There are two reasons of the fact. Lack of the appropriate vocabularies - presented ontologies remain still under development. Another reason is that we are not able to represent advanced mathematical structures of the theory in RDF and OWL (Skulimowski, 2010). We also want to notice that automatic creation of RDF links for scientific publications in physics seems to be out of scope in this domain. The links has to be created manually by the research community. Taking this into account we are developing a web tool which support the creation. The tool is shortly presented in the paper.

Future work should focus on development of the presented ontologies. In particular, 'harmonization' between the ontologies and others ontologies (e.g. Semantic Publishing and Referencing (SPAR) ontologies and SWAN Scientific Discourse Ontology may be required to remove overlap and conflicts. Moreover, it would be beneficial to provide general hints on how to create Linked Data for the case of research articles on physics. Thanks to such hints creation of Linked Data for a research articles could become a part of scientific activity and even a part of an editorial process in the future. The future development of the LYR tool should focus on increasing support during the creation of RDF links, for example automatic generation of URIs or suggesting links between resources. Moreover, visualization possibilities of RDF links stored in LYR should be extended. In future, care should be also taken to ensure consistency of RDF data created using the tool.

REFERENCES

Aberer, K., Boyarsky, A., Cudré-Mauroux, P., Demartini, G., and Ruchayskiy, O. (2011). Sciencewise: A web-based interactive semantic platform for scientific collaboration. In *Proceedings on the 10th International*

Semantic Web Conference (ISWC2011), Bonn, Germany.

Berners-Lee, T. (2006). *Linked data - design issues*. Retrieved January 12, 2013, from http://www.w3.org/DesignIssues/LinkedData.html.

Ciccarese, P., Wub, E., Wongb, G., Ocana, M., Kinoshita, J., Ruttenberg, A., and Clark, T. (2008). The SWAN biomedical discourse ontology. *Journal of Biomedical Informatics*, 41(5):739–751.

Dodds, L. and Davis, I. (2011). *Linked Data Patterns: A pattern catalogue for modelling, publishing, and consuming Linked Data*. Retrieved February 7, 2013, from http://patterns.dataincubator.org/book/linked-data-patterns.pdf.

Glasman-Deal, H. (2010). *A Guide for Non-Native Speakers of English*. London: Imperial College Press.

Goble, C. and Roure, D. D. (2007). myExperiment: Social networking for workflow-using e-scientists. In *Proceedings of the 2nd workshop on Workflows in support of large-scale science, Monterey, California, USA, ACM*, pages 1–2.

Heath, T. and Bizer, C. (2011). *Linked Data: Evolving the Web into a Global Data Space*. Morgan & Claypool Publishers.

Hepp, M. (2008). Goodrelations: An ontology for describing products and services offers on the web. In *Proceedings of the 16th International Conference on Knowledge Engineering and Knowledge Management, Acitrezza, Italy*.

Kauppinen, T., Baglatzi, A., and Keler, C. (2013). Linked science: Interconnecting scientific assets. In *Terence Critchlow and Kerstin Kleese-Van Dam (Eds.): Data Intensive Science*. CRC Press, USA.

Sheridan, J. and Tennison, J. (2010). Linking UK government data. In *Proceedings of the Linked Data on the Web Workshop (LDOW2010)*.

Skulimowski, M. (2010). An OWL ontology for quantum mechanics. In *Proceedings of the 7th International Workshop on OWL: Experiences and Directions (OWLED 2010), San Francisco*. Retrieved February 18, 2013, http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-614/.

Smrz, P. and Dytrych, J. (2011). Towards new scholarly communication: A case study of the 4A framework. In *Proceedings of the Workshop on the Semantic Publishing (SePublica)*. Retrieved February 18, 2013, http://ceur-ws.org/Vol-721/paper-07.pdf.

Sren Auer, J. L. and Hellmann, S. (2009). Linkedgeodata - adding a spatial dimension to the web of data. In *Proceedings of International Semantic Web Conference*.

Wiljes, C. and Cimiano, P. (2012). Linked data for the natural sciences: Two use cases in chemistry and biology. In *Proceedings of the Workshop on the Semantic Publishing (SePublica)*. Retrieved February 18, 2013, http://ceur-ws.org/Vol-903/paper-07.pdf.