



Data Knowledge Base for HENP Research Results and Interim Findings

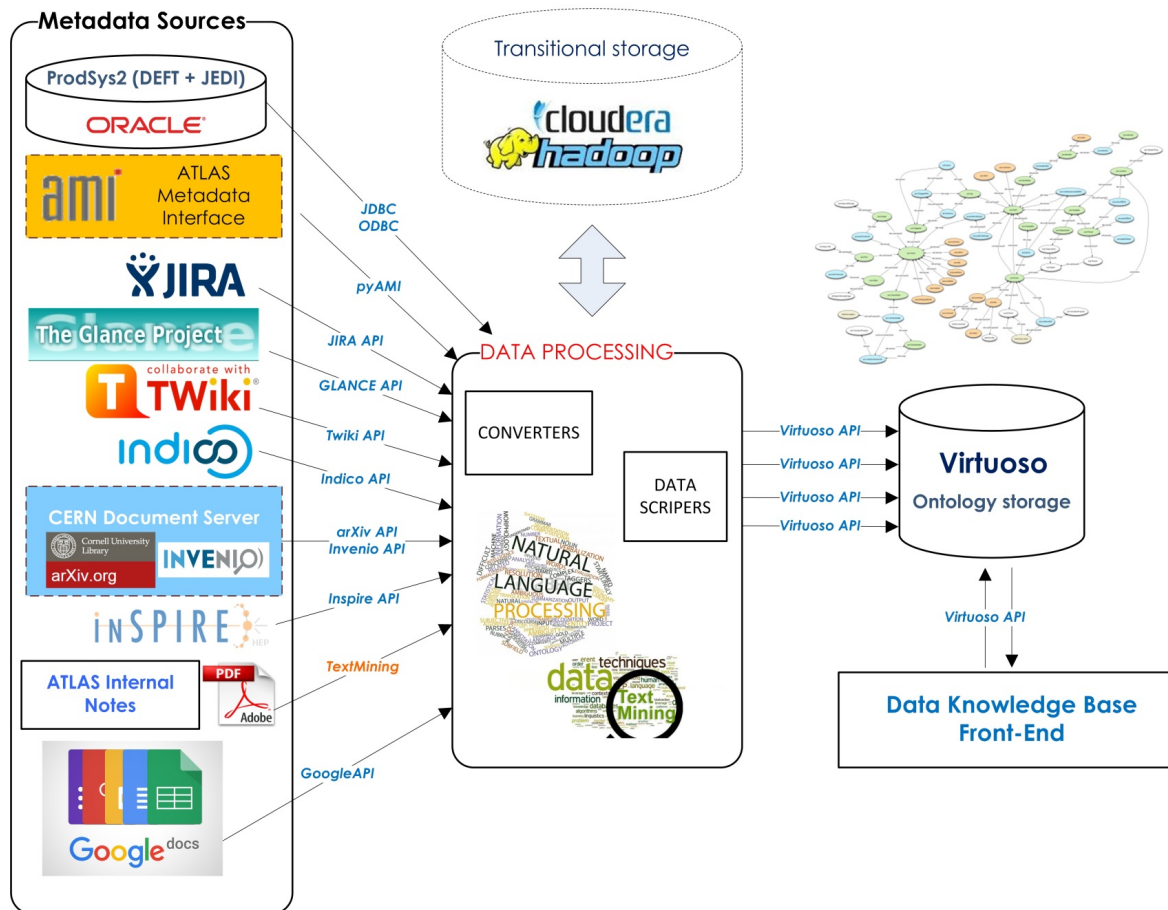
Grigorieva Maria (NRC KI)

Outline

- Previous architecture prototypes
 - Architecture Prototype v.1 (October 2016)
 - Architecture Prototype v.2 (February 2017)
- Ontological modelling
- Metadata extraction from PDF Documents
- Data Integration methods
- Current architecture prototype v.3 (April 2017)
- Dataflow (Apache Kafka) by Marina Golosova
- Near term plans
- Questions



DKB architecture prototype v.1



Ontology storage – OpenLink Virtuoso:

- Developed first prototype of the ontology for ATLAS Data Analysis.
- Virtuoso ontology storage installed in Tomsk Polytechnic University

Transitional Hadoop Storage in Kurchatov Institute

- Production System metadata [datasets] was exported from Oracle DB and imported to Hadoop Storage

ATLAS Internal Notes processing:

- Developed PDF Analyzer tool, based on PDFMiner, which extracts dataset names from full texts of ATLAS Internal Notes

Data Processing:

- Developed tools, converting the metadata from GLANCE, CERN Document Server and Production System [datasets] to TTL format for Virtuoso storage
- Metadata consolidated in Virtuoso storage





General Idea: use Semantic Web for Metadata Integration

- “*The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation.*” // Berners-Lee, *Scientific American*, May 2001
- Semantic Web consists primarily of three technical standards:
 - **RDF (Resource Description Framework)**
 - **SPARQL (SPARQL Protocol and RDF Query Language)**
 - **OWL (Web Ontology Language)**
- RDF Statements are expressed in a “**triples**” <subject, predicate, object>.
- The entire universe can be described by triples because together, triples comprise a **graph**.
- A graph can be linked to one or many other graphs on the World Wide Web and these graphs are a fundamental part of the Semantic Web.
- **Knowledge-oriented systems:** reasoning engines can be used to reason against assertions that have been made to infer new meaning, to find relationships and meaning far beyond the scope of the data, managed isolated.

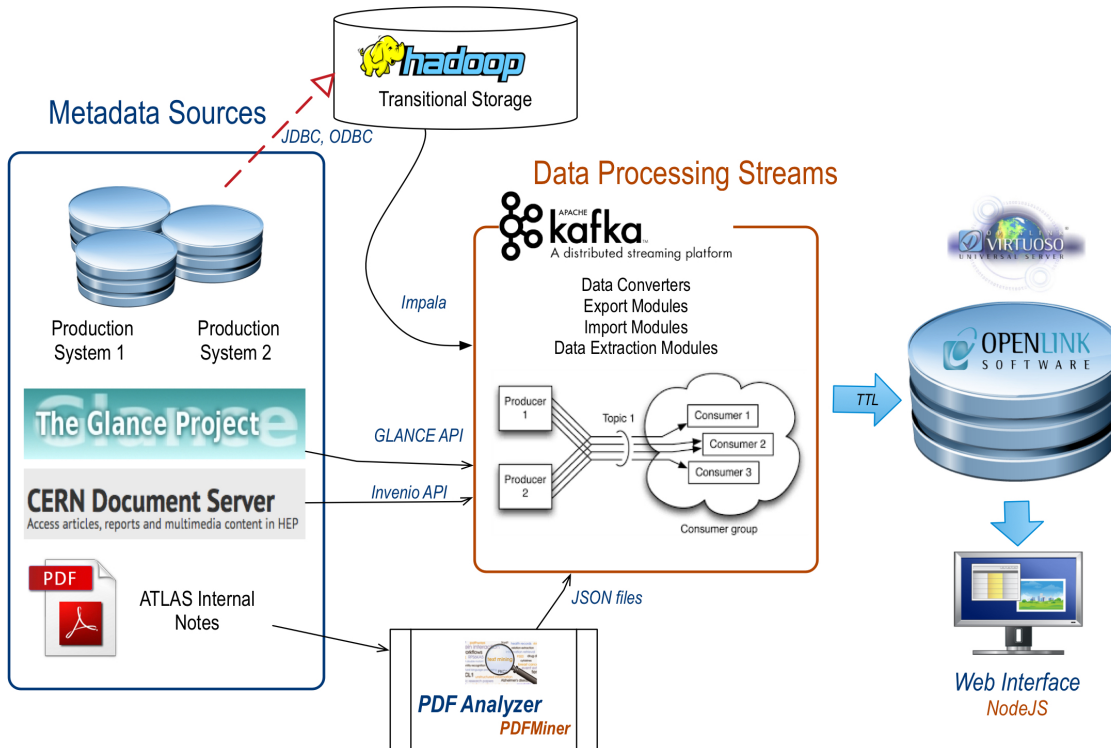




DKB architecture prototype v.2

Data Processing:

- Implemented Metadata Integration Chain using data streaming based on Apache Kafka to automate data processing workflows
- Executed test data flow which export metadata from GLANCE and CDS, and import integrated metadata in TTL format in Virtuoso Storage



Enhanced PDF Analyzer functionality:

- Extract metadata from PDF Tables

Enhanced ontological model with metadata from AMI:

- Data taking periods
- Run Numbers
- Projects
- Campaigns





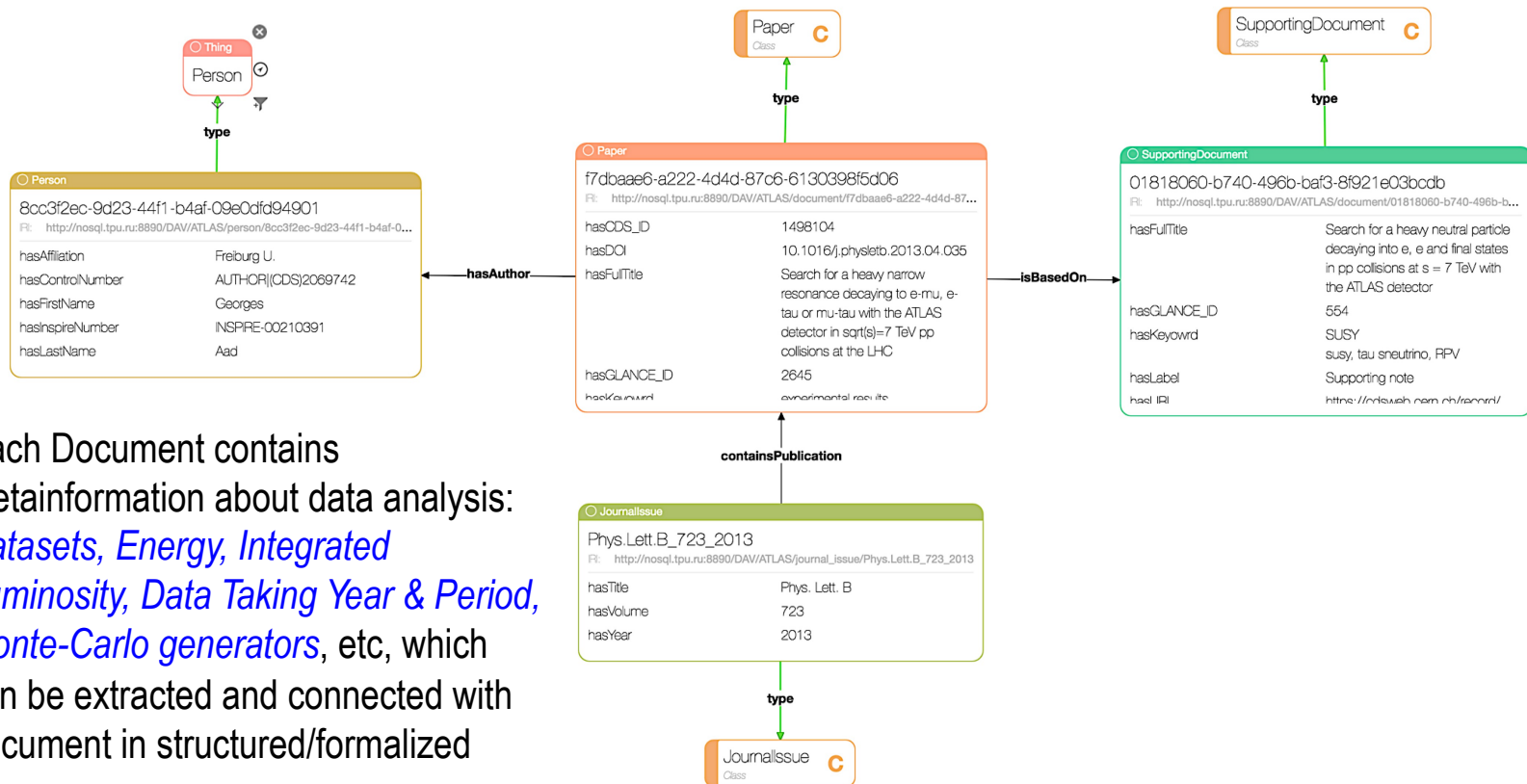
Architecture prototype v2: Tasks assignments

- **Data Processing modules for extracting/processing/converting metadata from GLANCE, CDS, ProdSys**
 - Marina Golosova (NRC KI)
 - Maria Grigorieva (NRC KI)
- **Kafka Streams for automated data processing**
 - Marina Golosova (NRC KI)
- **ATLAS Internal Notes PDF Analyzer**
 - Vasily Aulov (NRC KI)
- **Ontological model of data analysis in HEP**
 - Maria Grigorieva (NRC KI)
 - Maxim Gubin (TPU)
- **Web Interface**
 - Maria Grigorieva (NRC KI)
- **Virtuoso Server administration**
 - Alexander Alexeev (TPU)
- **Hadoop Transitional Storage administration**
 - Marina Golosova (NRC KI)



Ontological Modeling of ATLAS Data analysis metadata

- **Documents** can be of different types.
- **Scientific Paper** is based on **Supporting Documents**.
- Document's inheritance is provided by "**isBasedOn**" Object Attribute.

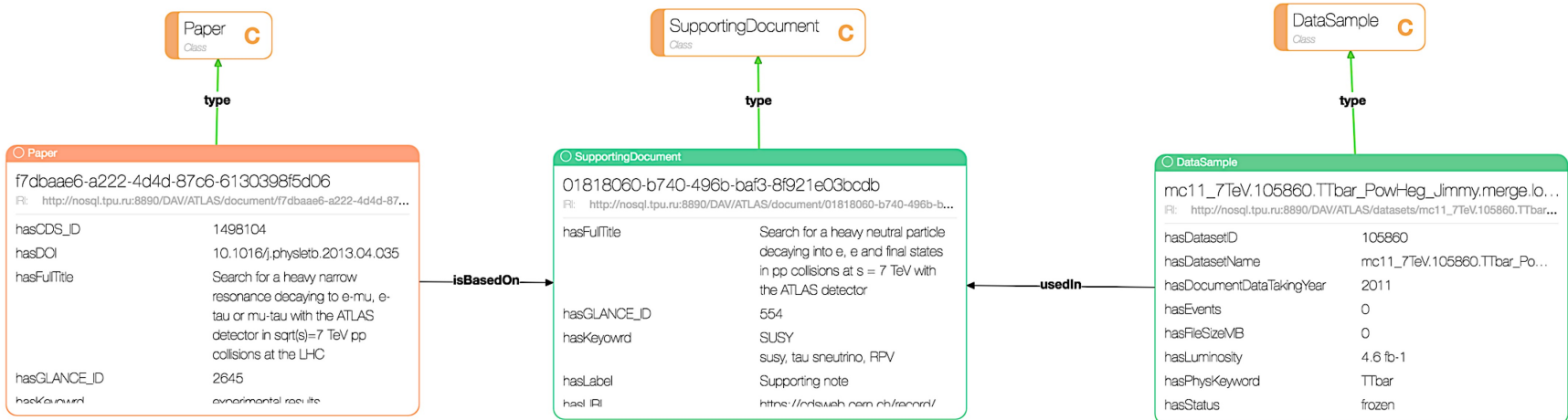


- Each Document contains metainformation about data analysis: *Datasets, Energy, Integrated Luminosity, Data Taking Year & Period, Monte-Carlo generators*, etc, which can be extracted and connected with document in structured/formalized view.



Ontological Modeling of ATLAS Data analysis metadata

- **DataSamples** and **Documents** are connected by “**usedIn**”/“**referTo**” attributes.
- DataSample attributes are taken from ProductionSystem database:
 - hasDatasetID
 - hasStatus
 - hasEvents
 - hasTimestamp
 -
- In architecture prototypes v1 and v2 dataset’s detailed metadata are consolidated in Virtuoso RDF-Storage.



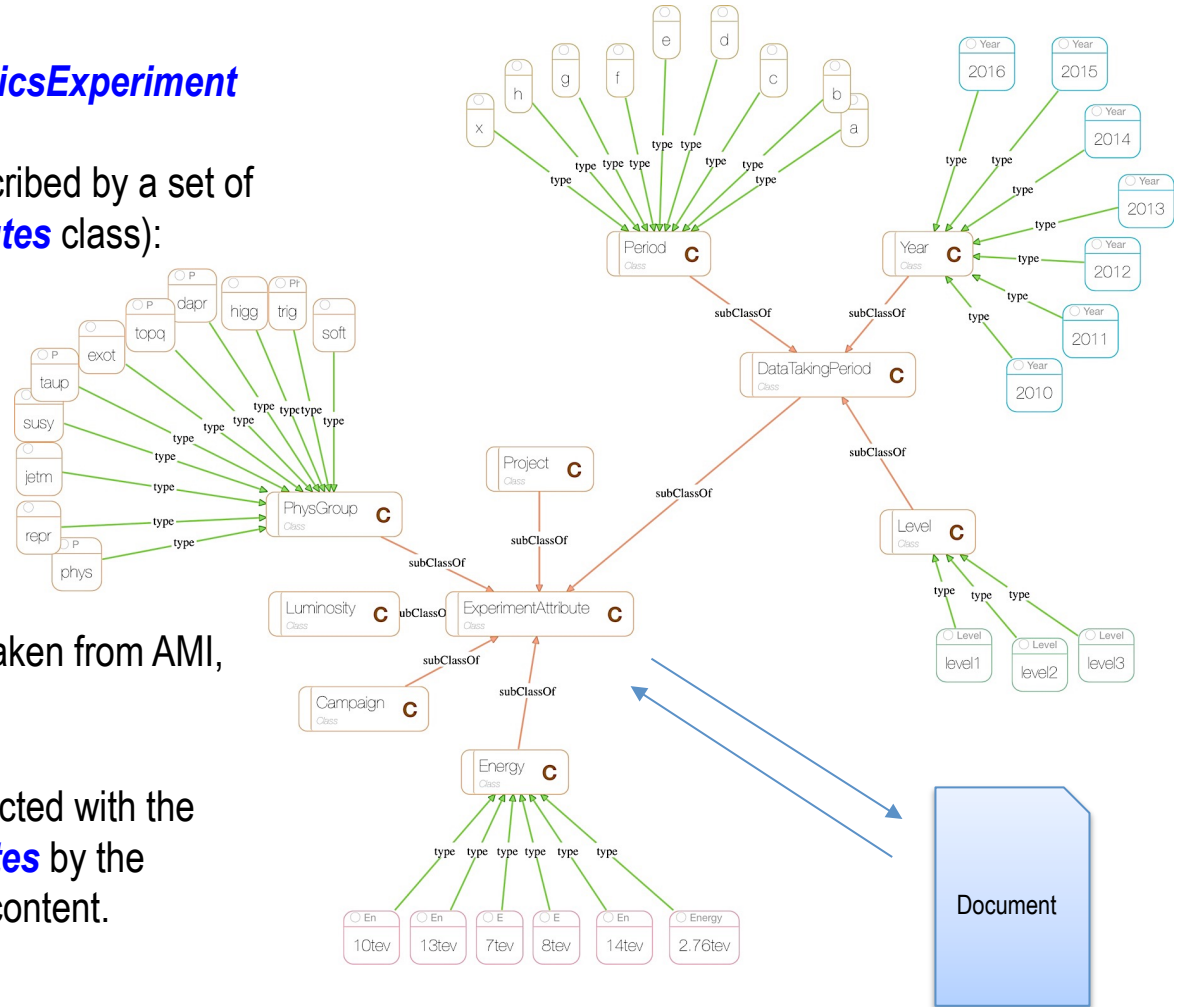
Ontological Modeling of ATLAS Data analysis metadata

- Data Analysis in ATLAS = **PhysicsExperiment** class
- Each PhysicsExperiment is described by a set of parameters (**ExperimentAttributes** class):

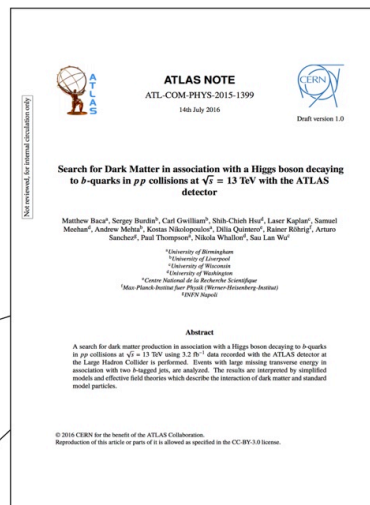
- Project (ex: mc10_7TeV)
- Campaign (ex: mc11a)
- Energy (ex: 10TeV)
- Integrated Luminosity
- Physics Group (SUSY, HIGG,...)
- Data Taking Period (ex: 2010_A1_1)
- *Other parameters are not defined yet*

- **ExperimentAttributes** can be taken from AMI, Production System, Twiki pages.

- Each **Document** must be connected with the appropriate **ExperimentAttributes** by the parameters, extracted from the content.



Document Contents



hasContent

hasContent

hasContent

hasContent

hasContent

GRL used in this analysis is DATA15_13TeV.PERIODALLYEAR_DETSTATUS-v73-PRO19-08_DQDEFFECTS-00-01-02_PHYS_STANDARDGRL_ALL_GOOD_25NS.XML

Abstract

A search for dark matter production in association with a Higgs boson decaying to b -quarks in pp collisions at $\sqrt{s} = 13$ TeV using 3.2 fb^{-1} data recorded with the ATLAS detector at the Large Hadron Collider is performed. Events with large missing transverse energy in association with two b -tagged jets, are analyzed. The results are interpreted by simplified models and effective field theories which describe the interaction of dark matter and standard model particles.

L Monte Carlo Samples

```

PYTHIA 6.423
mc10_7TeV.105009.J0_pythia_jetjet.merge.NTUP_BTAG.e574_s934_s946_r1653_r1700_p370
...
mc10_7TeV.105016.J7_pythia_jetjet.merge.NTUP_BTAG.e574_s934_s946_r1653_r1700_p370

PYTHIA 6.423+PILEUP
mc10_7TeV.105009.J0_pythia_jetjet.merge.NTUP_BTAG.e574_s934_s946_r1833_r1700_p370
...
mc10_7TeV.105016.J7_pythia_jetjet.merge.NTUP_BTAG.e574_s934_s946_r1833_r1700_p370
    
```

Real datasets

DSID	Sample Name	Tag
341100	Pythia8EvtGen_A14NNPDF23LO_WlvH125_bb	e3885_s2608_s2183_r6869_r6282_p2419
341101	Pythia8EvtGen_A14NNPDF23LO_ZvvH125_bb	e3885_s2608_s2183_r6869_r6282_p2419
341102	Pythia8EvtGen_A14NNPDF23LO_ZlIH125_bb	e3885_s2608_s2183_r6869_r6282_p2419

Table 19: Monte Carlo samples used as baseline for Standard Model $VH(\rightarrow bb)$.

DSID	Sample Name	Tag
00279598	physics_Main	f628_m1497_p2425
00279685	physics_Main	f628_m1497_p2425
00279764	physics_Main	f628_m1497_p2425
00279813	physics_Main	f628_m1497_p2425
00279867	physics_Main	f628_m1497_p2425
00279928	physics_Main	f628_m1497_p2425
00279932	physics_Main	f629_m1504_p2425
00279984	physics_Main	f629_m1504_p2425
00280231	physics_Main	f630_m1504_p2425
00280319	physics_Main	f629_m1504_p2425
00280368	physics_Main	f629_m1504_p2425
00280423	physics_Main	f629_m1504_p2425
00280464	physics_Main	f629_m1504_p2425
00280500	physics_Main	f631_m1504_p2425
00280520	physics_Main	f632_m1504_p2425
00280614	physics_Main	f629_m1504_p2425



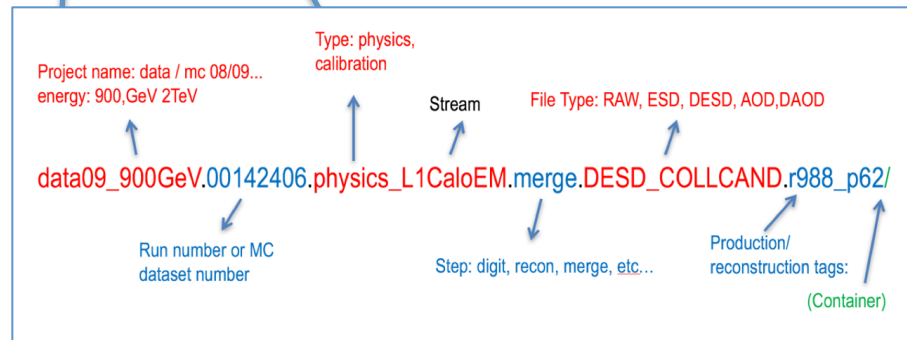
Datasets extraction from PDF documents

```

internal circulation only
743 A.1 Data PDF
744 Egamma stream
745 data11_7TeV.00178044.physics.Egamma.merge.DAOD_2LHSG2.f354_m765_p600/
746 data11_7TeV.00178047.physics.Egamma.merge.DAOD_2LHSG2.f351_m765_p600/
747 data11_7TeV.00178109.physics.Egamma.merge.DAOD_2LHSG2.f354_m765_p600/
748 data11_7TeV.00179710.physics.Egamma.merge.DAOD_2LHSG2.f361_m796_p600/
    
```

Regular expressions are constructed according to ATLAS dataset nomenclature. These expressions are used to extract the dataset names from the text.

PDFMiner is a tool for extracting information from PDF documents.



GUI interface:

- Edit dataset names
- Delete datasets from the list
- Export resulted list to JSON file

```

realdata
spaces data11_7TeV.00178044.physics.Egamma.merge.DAOD_2LHSG2.f354_m765_p600
spaces data11_7TeV.00178044.physics.Muons.merge.DAOD_2LHSG2.f354_m765_p600
spaces data11_7TeV.00178047.physics.Egamma.merge.DAOD_2LHSG2.f351_m765_p600
    
```

JSON

```

{ "content": { "real_datasets": [ "data11_7TeV.
00178044.physics_Egamma.merge.DAOD_2LHSG2.f354_m765_p600",
"data11_7TeV.00178047.physics_Egamma.merge.DAOD_2LHSG2.f351_m765_p600",
"data11_7TeV.00178109.physics_Egamma.merge.DAOD_2LHSG2.f354_m765_p600",
"data11_7TeV.00179710.physics_Egamma.merge.DAOD_2LHSG2.f361_m796_p600",
"data11_7TeV.00179725.physics_Egamma.merge.DAOD_2LHSG2.f361_m796_p600",
.....
    
```



Extraction of data from tables

Signal Point		Run Number	Cross Section (LO) [pb]	Signal Point		Run Number	Cross Section (LO) [pb]
$M(\tilde{g})$ [GeV]	$M(\tilde{\chi}_1^0)$ [GeV]			$M(\tilde{g})$ [GeV]	$M(\tilde{\chi}_1^0)$ [GeV]		
400	50	123078	6.00	900	50	138568	6.19×10^{-3}
400	75	123079	5.95	900	100	138569	6.14×10^{-3}
400	100	123080	6.00	900	150	138570	6.10×10^{-3}
400	125	123081	6.02	900	200	138571	6.08×10^{-3}
400	150	118430	6.03	900	300	138572	5.92×10^{-3}

Table 26: GGM signal samples. Each signal point is defined by the gluino and lightest neutralino mass. The run number and LO cross section is given.

```
<?xml version="1.0" encoding="utf-8" ?>
<pages>
  <page id="1" bbox="0.000,0.000,612.000,792.000" rotate="0">
    <textbox id="0" bbox="229.080,675.825,366.146,691.074">
      <textline bbox="229.080,675.825,366.146,691.074">
        <text font="MGTNRE+CMSSBX10" bbox="229.080,675.825,244.226,691.074" size="15.249">A</text>
        <text font="MGTNRE+CMSSBX10" bbox="242.273,675.825,257.419,691.074" size="15.249">T</text>
        <text font="MGTNRE+CMSSBX10" bbox="257.386,675.825,269.391,691.074" size="15.249">L</text>
        <text font="MGTNRE+CMSSBX10" bbox="269.396,675.825,284.541,691.074" size="15.249">A</text>
        <text font="MGTNRE+CMSSBX10" bbox="284.509,675.825,297.134,691.074" size="15.249">S</text>
      </textline>
    </textbox>
    <figure name="R5" bbox="70.920,642.800,155.880,750.800">
      <image width="84" height="108" />
    </figure>
    <layout>
      <textgroup bbox="73.080,185.033,522.095,691.074">
        <textgroup bbox="73.080,380.838,522.095,691.074">
          </layout>
        </textgroup>
      </textgroup>
    </layout>
  </page>
```

Tables are found by looking for their descriptions in text. After finding the description, the page containing the table is extracted into XML format and the table is reconstructed using the information from it.

```
"table_26": [
  "Table 26: GGM signal samples. Each signal point is defined by the gluino and lightest neutralino mass. The run number and LO cross section is given."
  [
    [
      "Signal",
      "Point",
      "Run Number",
      "Cross Section (LO)",
      "Signal",
      "Point",
      "Run Number",
      "Cross Section (LO)"
    ],
    [
      "M( \u02dcg)",
      "M( \u02dc\u03c7\u2081\u2070)",
      "EMPTY",
      "[pb]",
      "M( \u02dcg)",
      "M( \u02dc\u03c7\u2081\u2070)",
      "EMPTY",
      "[pb]"
    ],
    [
      "[GeV]",
      "[GeV]",
      "EMPTY",
      "EMPTY",
      "[GeV]",
      "[GeV]",
      "EMPTY",
      "EMPTY"
    ],
    [
      "400",
      "50",
      "123078",
      "6.00",
      "900",
      "50",
      "138568",
      "6.19\u22123"
    ],
    [
      "400",
      "75",
      "123079",
      "5.95",
      "900",
      "100",
      "138569",
      "6.14\u22123"
    ],
    [
      "400",
      "100",
      "123080",
      "6.00",
      "900",
      "150",
      "138570",
      "6.10\u22123"
    ],
    [
      "400",
      "125",
      "123081",
      "6.02",
      "900",
      "200",
      "138571",
      "6.08\u22123"
    ],
    [
      "400",
      "150",
      "118430",
      "6.03",
      "900",
      "300",
      "138572",
      "5.92\u22123"
    ]
  ]
]
```



Virtuoso Server = <http://nosql.tpu.ru:8890/>

SCHEMA GRAPH:

<http://nosql.tpu.ru:8890/DAV/home/dba/ATLAS>



RESOURCE GRAPH:

<http://nosql.tpu.ru:8890/DAV/ATLAS>

Subject	Predicate	Object
Document/CERN-EP-2016-181	hasType	Paper
Document/CERN-EP-2016-182	hasTitle	Search for dark matter in association with a Higgs boson decaying to b-quarks in pp collisions at sv=13 TeV with the ATLAS detector
Document/CERN-EP-2016-183	hasArXiv	arXiv:1609.04572
Document/CERN-EP-2016-184	hasKeyword	exotics
Document/CERN-EP-2016-185	isBasedOn	Document/ATL-COM-PHYS-2015-1399
Document/ATL-COM-PHYS-2015-1399	hasType	SupportingDocument
Document/ATL-COM-PHYS-2015-1400	hasTitle	Search for dark matter in association with a Higgs boson decaying to b-quarks in pp collisions at sv=13 TeV with the ATLAS detector
Document/ATL-COM-PHYS-2015-1401	useGRL	data15_13TeV.periodAllYear_DetStatus-v73-pro19-08_DQDefects-00-01-02_PHYS_StandardGRL_All_Good_25ns.xml
Document/ATL-COM-PHYS-2015-1402	hasEnergy	13TeV
Document/ATL-COM-PHYS-2015-1403	hasPublicationYear	2016
Document/ATL-COM-PHYS-2015-1404	hasContent	
Document/ATL-COM-PHYS-2015-1405		Content/ATL-COM-PHYS-2015-1399_Table_19
Content/ATL-COM-PHYS-2015-1399_Table_19	hasDescription	Monte Carlo samples used as baseline for Standard Model VH(→ bb)
Content/ATL-COM-PHYS-2015-1399_Table_20	mentionsDataSample	DataSample/ATL-COM-PHYS-2015-1399_341100
Content/ATL-COM-PHYS-2015-1399_Table_21	mentionsDataSample	DataSample/ATL-COM-PHYS-2015-1399_341101
Content/ATL-COM-PHYS-2015-1399_Table_22	mentionsDataSample	DataSample/ATL-COM-PHYS-2015-1399_341102
DataSample/ATL-COM-PHYS-2015-1399_341100	hasType	MC
DataSample/ATL-COM-PHYS-2015-1399_341100	hasProject	MC15_13TeV
DataSample/ATL-COM-PHYS-2015-1399_341100	hasDataSampleID	341100
DataSample/ATL-COM-PHYS-2015-1399_341100	hasSampleName	Pythia8EvtGen_A14NNPDF23LO_WlvH125_bb
DataSample/ATL-COM-PHYS-2015-1399_341100	hasTag	e3885_s2608_s2183_r6869_r6282_p2419



SPARQL Endpoint:
<http://nosql.tpu.ru:8890/sparql>



Classes

Search for...

- (1)
- AllDisjointClasses (1)
- array-of-QuadMap (3)
- array-of-QuadMapATable (2)
- array-of-QuadMapColumn (8)
- array-of-QuadMapFormat (98)
- array-of-string (2)
- ATLASMember (0)
- DataSample (162686)
- DataSample (43643)
- DataSample (8)
- Document (0)
- Document (5)
- ExperimentAttribute (767)
- Individual (1)
- Male (1)
- NamedIndividual (661)
- Nothing (0)
- OnlineAccount (1)
- Ontology (5)
- Person (2)

Instances

Connected to 01818060-b740-496b-baf3-8f921e03bccdb

Search for...

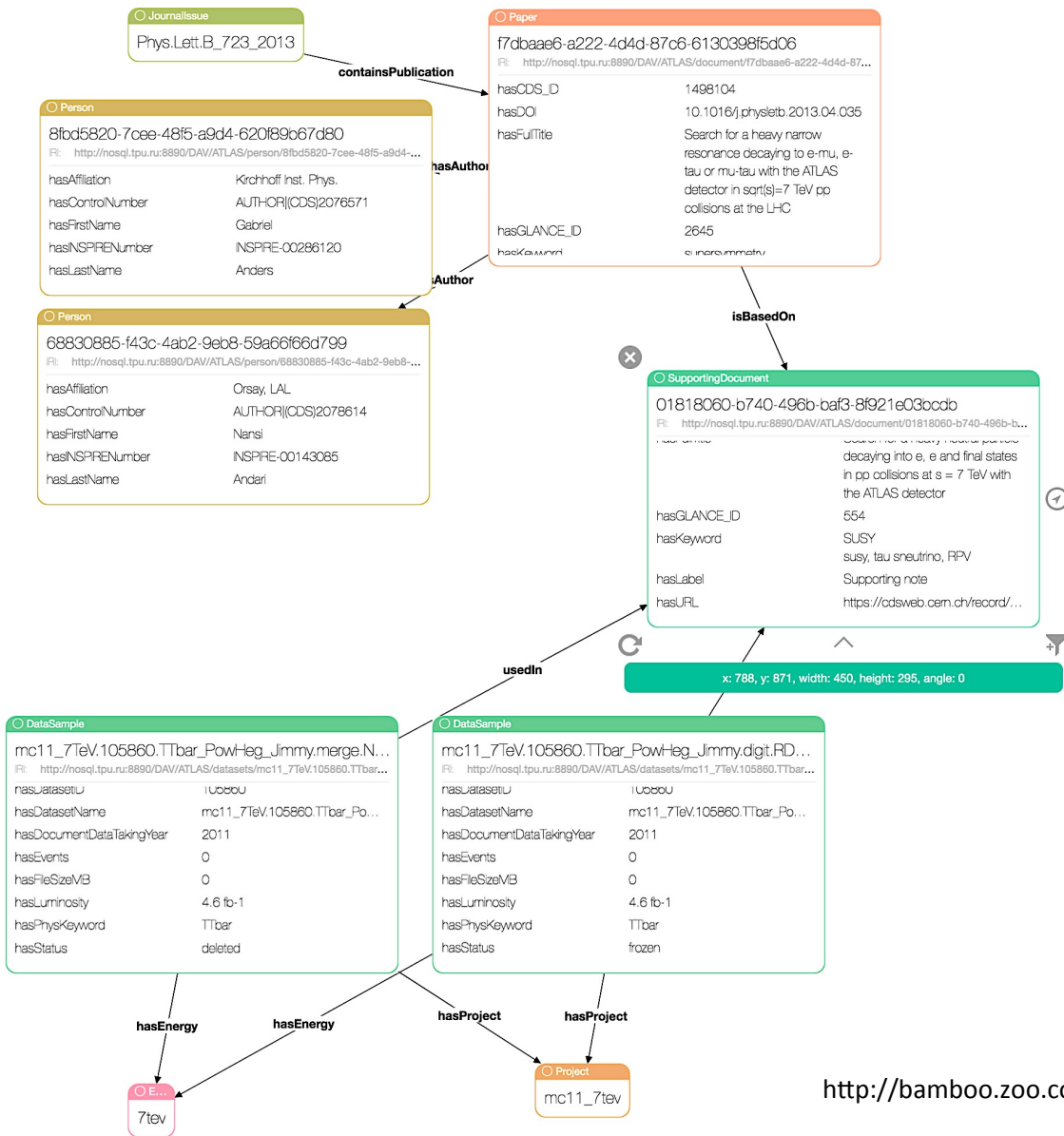
mc11_7TeV.105860.TTbar_PowHeg_Jimmy.merge.N...
owHeg_Jimmy.merge.NTUP_S
MWENU.e873_s1310_s1300_r
2730_r2780_p801_tid59130
8_00

mc11_7TeV.114612.SherpaW
5jetstomunu30GeV.recon.l
og.e931_s1310_s1300_r273
0_tid541782_00

mc11_7TeV.105860.TTbar_P
owHeg_Jimmy.simul.HITS.e
1198_a131_tid785061_00

mc11_7TeV.105860.TTbar_P
owHeg_Jimmy.merge.NTUP_S
USY01LEP.e873_s1310_s130
0_r3043_r2993_0935_tid76

Web Interface for Virtuoso Storage



Connections

Search for...

Switch all

Connected to 01818060-b740-496b-baf3-8f921e03bccdb

isbasedon 1

type 1

usedin 12896

Ontodia is a JavaScript library that allows to visualize, navigate and explore the data in the form of an interactive graph based on underlying data sources.

Data Integration Techniques

Data Consolidation captures data from multiple source systems and integrates it into a single persistent data store

- + Optimal performance and stability
- Requested data might be out of date because of the complexity of providing data synchronization with various data sources

Data Federation provides a single virtual view of one or more source data sources

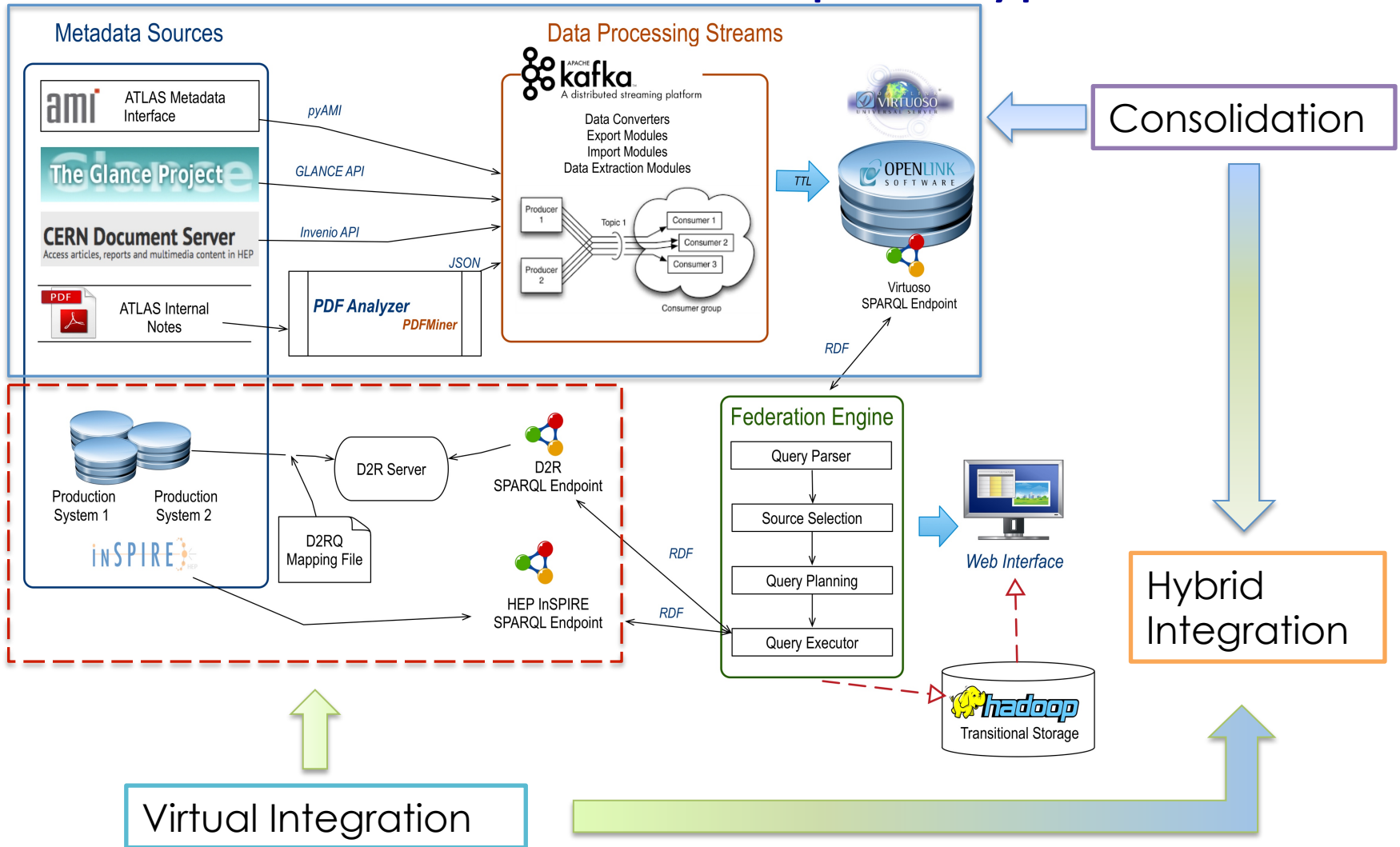
- + Federated requests always return the actual data – data integrity support remains on the data source side
- The performance of federated queries depends on the communication channels and the queries execution rate on the side of data sources

Hybrid Data Integration:

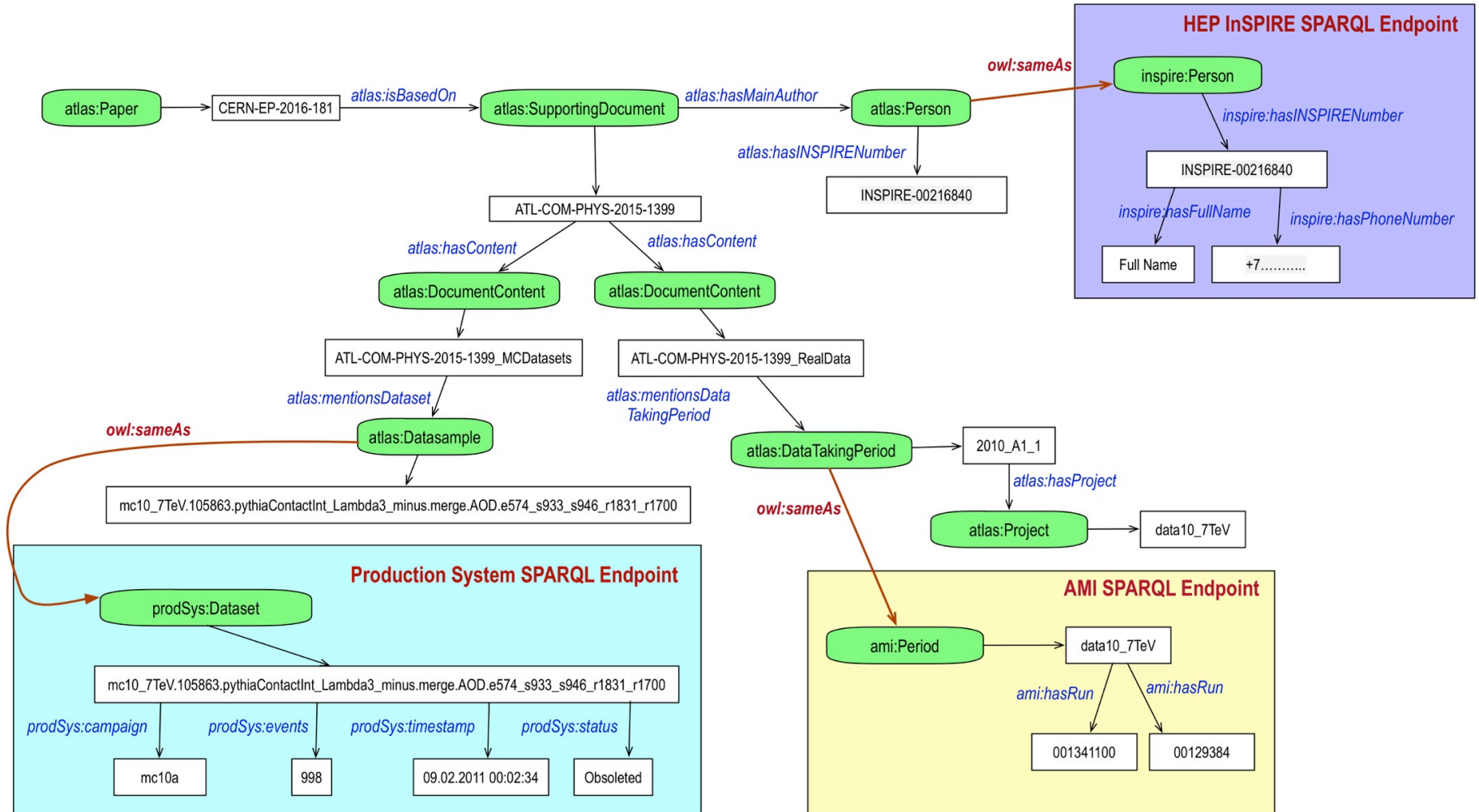
- Unchangeable data about global objects is consolidated:
 - Experiment Attributes
 - Links between documents
 - Documents general parameters from CDS
 - Metadata extracted from document's content
- Changing and auxiliary data is federated:
 - Detailed documents and authors metadata from HEP InSPIRE system
 - Dataset's detailed metadata from ProdSys/AMI/Rucio



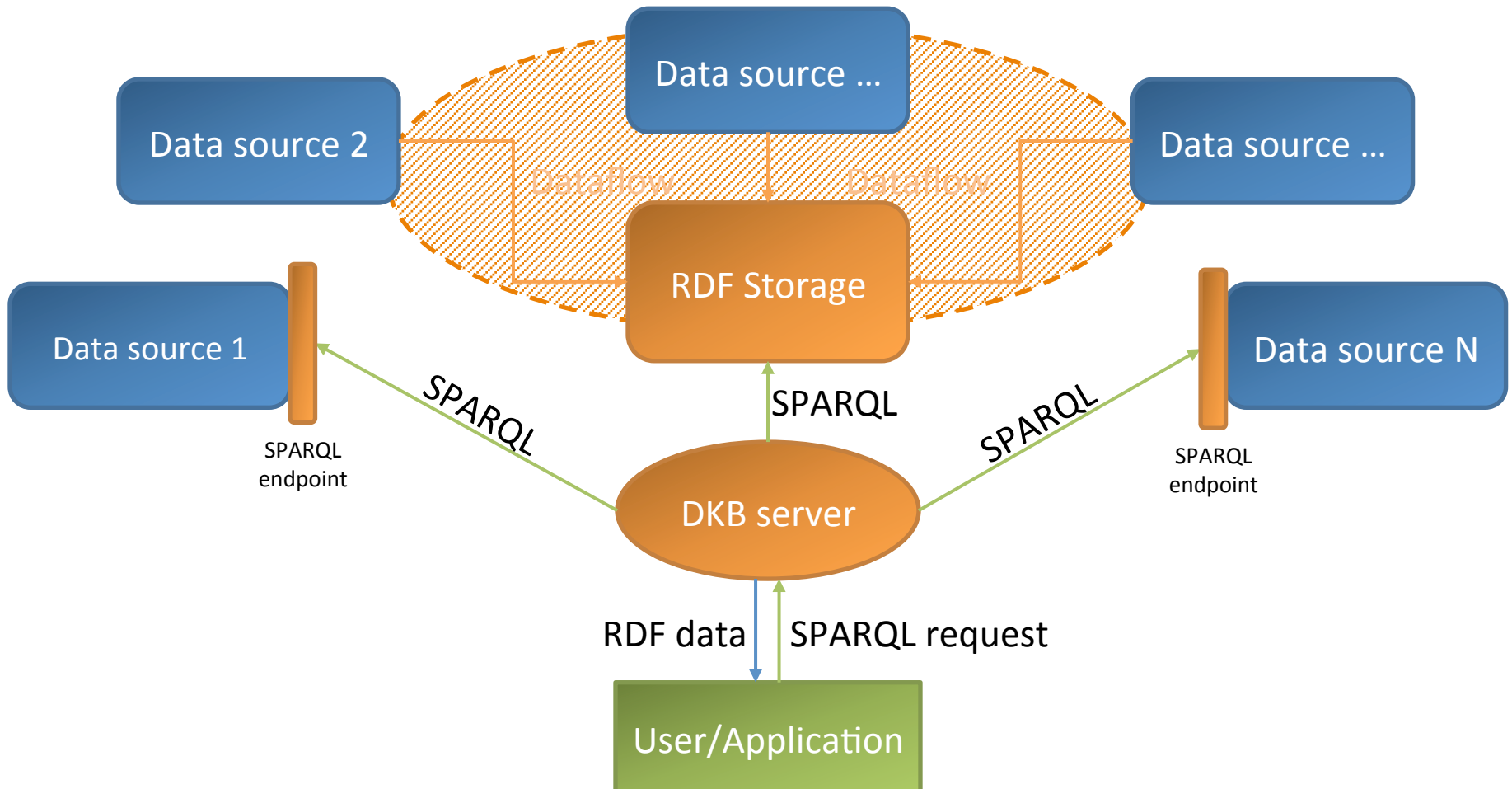
Current architecture prototype v.3



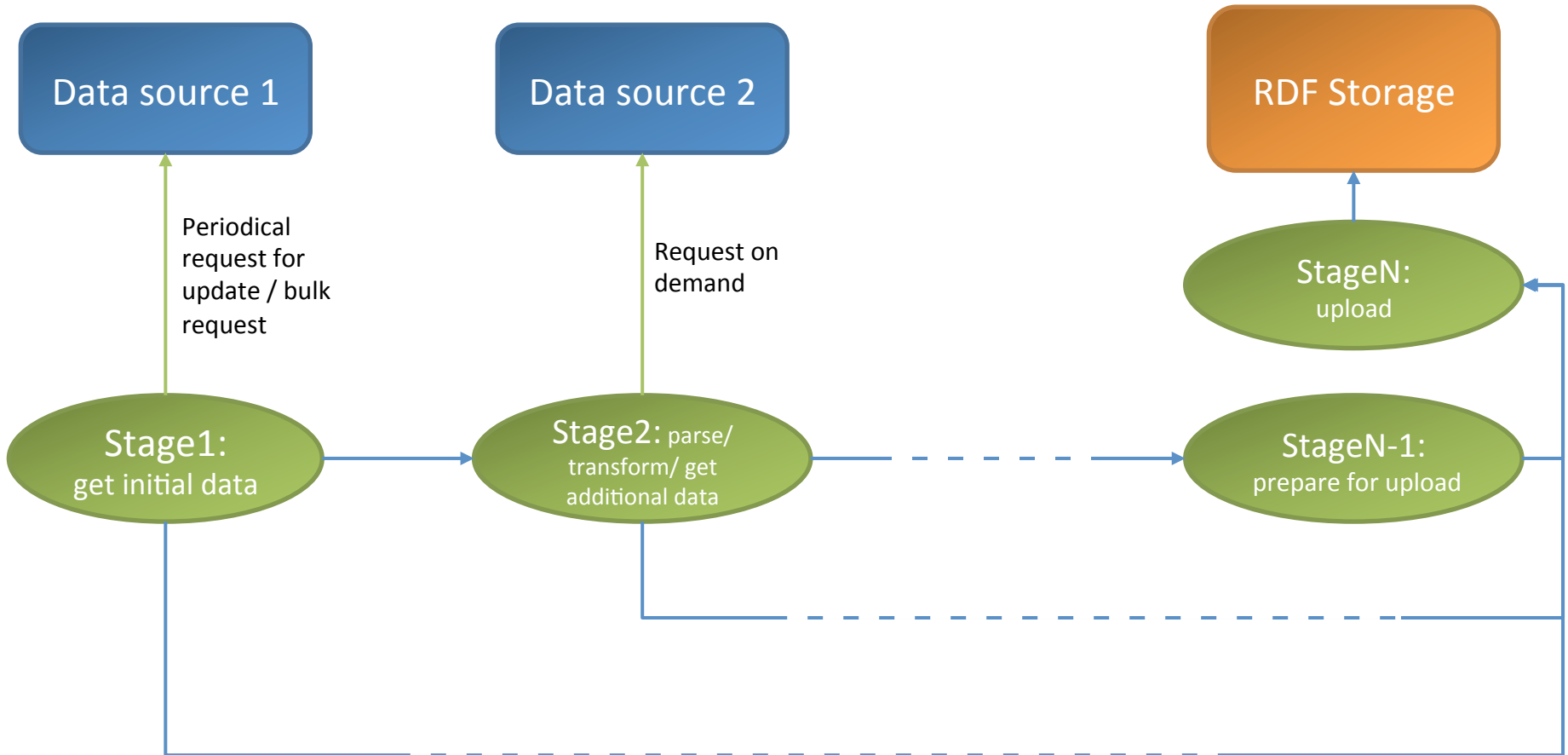
Federated Requests Prototype



DKB schematic



Dataflow schematic



Dataflow tasks

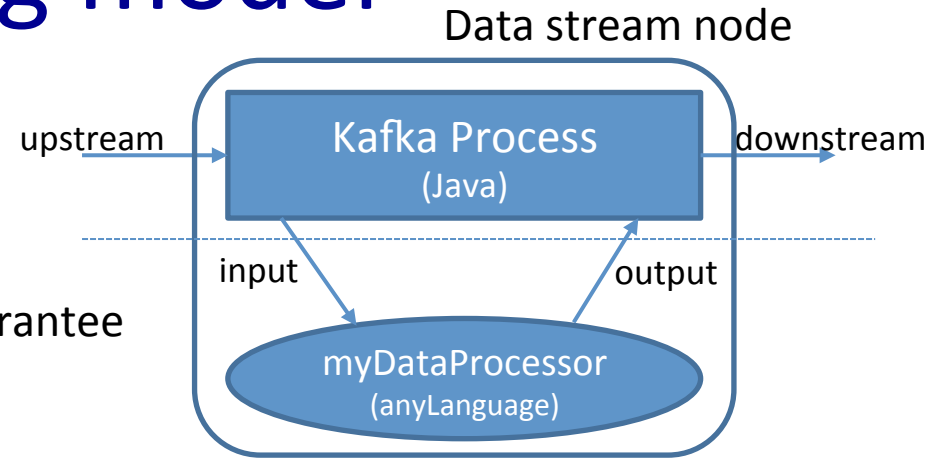
- ? Data transfer from “upstream” to “downstream” stages
- ? Obsolete data removal
- ? Data preservation between stages
- ? Guarantee of reprocessing on data/process update
- ? Processing delay
- ? Failure recovery



Streaming model



- ✓ **Message broker**
data preservation and delivery guarantee
- ✓ **Stream processing**
automation and delay minimizing;
no need to use additional processing systems (Spark, ...)
- ✓ **Distributed system**
scalability and reliability



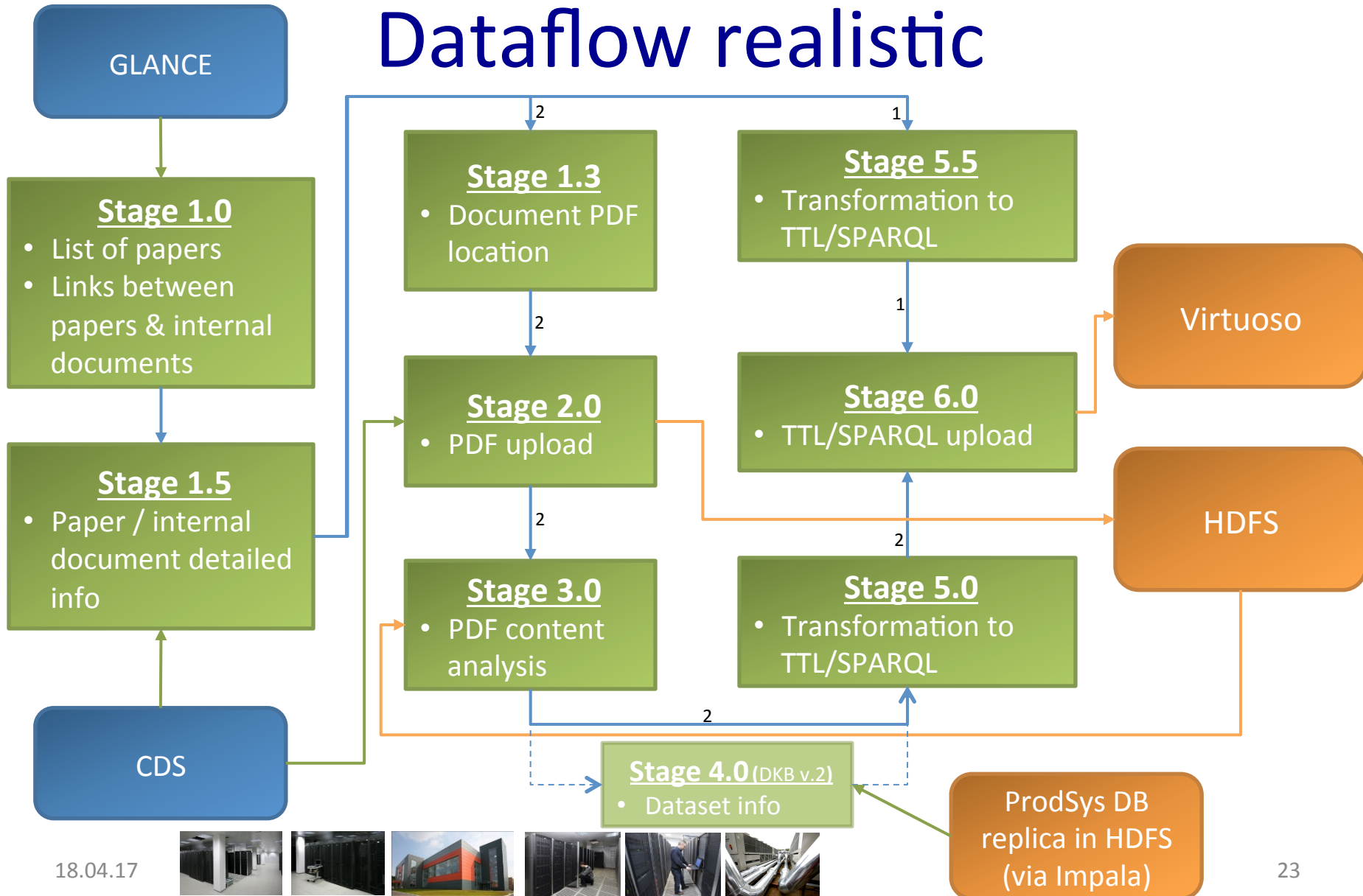
Developed: adapters to reuse existing data processing (extracting, transforming, uploading, ...) modules within Kafka-driven dataflow

Developed: management utility to run the dataflow stages

Formulated: basic rules to write dataflow modules (concerning data in/output, mainloop organization, error handling, ...)



Dataflow realistic



Near term plans

- Continue development of a global integration model of metadata, based on hybrid data integration approach
- Development of the ontological data models for various sources of ATLAS metadata
- Data schemes mismatch conflicts resolution
- Completion of the Kafka Streams dataflows automation components implementation
- Execution of the consolidation dataflow, based on metadata from GLANCE, CDS, PDF Analyzer
- The choice of technology and implementation of a SPARQL endpoint for ProdSys1&2 / Rucio / AMI (?), HEP InSPIRE / CDS (?)
- Enhancing semi-automatic PDF Analyzer functionality: providing better GUI, allowing users to correct the results of automatic data extraction
- **Execution of the federated SPARQL requests:**
 - get all datasets and their's attributes from ProdSys, which mentions in Supporting Documents for some scientific Paper
 - get all Documents, which refers to data analysis based on real data, taken from detector in 2014, period C
 - get all Documents, having physics group "SUSY", published in 2015 and refers to datasets, which status in ProdSys is "deleted"
 - get detailed information about **main** authors for Paper from InSPIRE and return titles of all theirs ATLAS publications, relates to project "mc10_7TeV"
 - get all Documents where dataset "data11_7TeV.00178044.physics_Egamma.merge.DAOD_2LHSG2.f354_m765_p600" is mentioned, with detailed metadata about this dataset from ProdSys, and find metadata about main authors of this document in InSPIRE



Questions:

- Which set of parameters is enough to [uniquely] identify data analysis (PhysicsExperiment)
 - Project (ex: mc10_7TeV)
 - Campaign (ex: mc11a)
 - Energy (ex: 10TeV)
 - Integrated Luminosity
 - Physics Group (SUSY, HIGG)
 - Data Taking Period (ex: 2010_A1_1)
- What are the preferred metadata sources for:
 - ATLAS Projects (full description – AMI, Twiki Pages, COMA) ?
 - ATLAS Campaigns (full description – AMI, Twiki Pages, COMA) ?
 - ATLAS Data Taking Periods (AMI, COMA) ?
 - Datasets (ProductionSystem/Rucio/AMI) ?
 - Authors & Publications (CDS/InSPIRE) ?



Questions:

- Metadata, which must be extracted from PDF Documents:
 - Data Taking Year
 - Projects (MC, Data)
 - Campaigns
 - Physics Groups
 - Good Run Lists
 - Software Release
 - The center-of-mass Energy
 - Integrated Luminosity
- Datasets:
 - MC
 - Full dataset names
 - Dataset ID
 - MC generators
 - AMI tags
 - Type (signal/background)
 - Real Data
 - Full dataset name
 - Data Taking period
 - Run Numbers
 - Triggers



Acknowledgements

- Many thanks for help and assistance:
 - Dmitry Golubkov
 - Alexei Klimentov
 - Mikhail Korotkov
 - Dimitry Krasnopevtsev
 - Eygene Ryabinkin
 - Anatoly Tuzovsky





ADDITIONAL SLIDES

Document Contents Analysis

Document Content

{JSON}

TTL



Abstract
 A search for dark matter production in association with a Higgs boson decaying to b-quarks in pp collisions at $\sqrt{s} = 13$ TeV using 3.2 fb⁻¹ data recorded with the ATLAS detector at the Large Hadron Collider is performed. Events with large missing transverse energy in association with two b-tagged jets, are analyzed. The results are interpreted by simplified models and effective field theories which describe the interaction of dark matter and standard model particles.

JSON

```
{ "luminosity": «3.2 fb-1»,  
  "energy": «13 TeV», }
```

TTL

```
Content/X :mentionsLuminosity «3.2 fb-1» .  
Content/X :mentionsEnergy «13 TeV»
```

hasContent

DSID	Sample Name	Tag
341100	Pythia8EvtGen_A14NNPDF23LO_WlVH125_bb	e3885_s2608_s2183_r6869_p282_p2419
341101	Pythia8EvtGen_A14NNPDF23LO_ZwH125_bb	e3885_s2608_s2183_r6869_p282_p2419
341102	Pythia8EvtGen_A14NNPDF23LO_ZBH125_bb	e3885_s2608_s2183_r6869_p282_p2419

Table 19: Monte Carlo samples used as baseline for Standard Model VH(→ bb).

JSON

```
{ «Table 19: Monte Carlo samples used as baseline for Standard Model  
VH(→ bb).»,  
  [  
    [  
      «DSID»,  
      «Sample Data»,  
      «Tag»,  
    ],  
    [  
      "341100",  
      "Pythia8EvtGen_A14NNPDF23LO_WlVH125_bb",  
      "e3885_s2608_s2183_r6869_p282_p2419"  
    ],  
    ...  
  ]  
}
```

hasContent

DSID	Sample Name	Tag
00279598	physics_Main	f628_m1497_p2425
00279685	physics_Main	f628_m1497_p2425
00279764	physics_Main	f628_m1497_p2425
00279813	physics_Main	f628_m1497_p2425
00279867	physics_Main	f628_m1497_p2425
00279928	physics_Main	f628_m1497_p2425
00279932	physics_Main	f629_m1504_p2425
00279984	physics_Main	f629_m1504_p2425
00280231	physics_Main	f630_m1504_p2425
00280319	physics_Main	f629_m1504_p2425
00280368	physics_Main	f629_m1504_p2425
00280423	physics_Main	f629_m1504_p2425
00280464	physics_Main	f629_m1504_p2425
00280500	physics_Main	f631_m1504_p2425
00280520	physics_Main	f632_m1504_p2425
00280614	physics_Main	f629_m1504_p2425

JSON

```
{ «Table 18: Data used as input in the CxAOD  
samples production version v14.»,  
  [  
    [  
      «DSID»,  
      «Sample Data»,  
      «Tag»,  
    ],  
    [  
      "00279598",  
      "physics_Main",  
      "f628_m1497_p2425"  
    ],  
    ...  
  ]  
}
```

TTL
 (?) Tabular data converting is under development

L Monte Carlo Samples

```
PYTHIA_6.423  
mc10.7TeV.105009.J0_pythia_jetjet.merge.NTUP_BTAG.e574_s934_s946.r1653.r1700.p370  
...  
mc10.7TeV.105016.J7_pythia_jetjet.merge.NTUP_BTAG.e574_s934_s946.r1653.r1700.p370  
  
PYTHIA_6.423+PileUP  
mc10.7TeV.105009.J0_pythia_jetjet.merge.NTUP_BTAG.e574_s934_s946.r1833.r1700.p370  
...  
mc10.7TeV.105016.J7_pythia_jetjet.merge.NTUP_BTAG.e574_s934_s946.r1833.r1700.p370
```

JSON

```
"datasets": {  
  "montecarlo": [  
    "mc10_7TeV.105009.J0_pythia_jetjet.merge.NTUP_BTAG.e574_s934",  
    "mc10_7TeV.105016.J7_pythia_jetjet.merge.NTUP_BTAG.e574_s934",  
    "mc10_7TeV.105009.J0_pythia_jetjet.merge.NTUP_BTAG.e574_s934",  
    "mc10_7TeV.105016.J7_pythia_jetjet.merge.NTUP_BTAG.e574_s934",  
    "mc10_7TeV.113066.J0_Pythia-EvtGen.merge.NTUP_BTAG.e635_s933",  
  ],  
}
```



Reconstruction of the real and MC datasets from document content

B Signal MC Samples

Table 21 shows a full list of signal MC samples of MC11a used in the council conf note. For the paper, MC11b is used. The difference on AMI tag is r2920_r2900.p756 instead of r2730_r2700.p756.

Table 21: List of signal MC samples used (MC11b production).

	Process	Dataset ID	Generator	Filter	AMI tag
$m_H=100$ GeV	$gg \rightarrow H$	116866	PowHeg	-	e873_s1310_s1300_r2920_r2900_p756
	VBF	125170	PowHeg	-	e893_s1310_s1300_r2920_r2900_p756
	WH	125329	Pythia	-	e825_s1310_s1300_r2920_r2900_p756
	ZH	125489	Pythia	-	e825_s1310_s1300_r2920_r2900_p756
	$t\bar{t}H$	116064	Pythia	PhotonFilter	e893_s1310_s1300_r2920_r2900_p756
$m_H=105$ GeV	$gg \rightarrow H$	116867	PowHeg	-	e873_s1310_s1300_r2920_r2900_p756
	VBF	125171	PowHeg	-	e893_s1310_s1300_r2920_r2900_p756
	WH	125330	Pythia	-	e825_s1310_s1300_r2920_r2900_p756
	ZH	125490	Pythia	-	e825_s1310_s1300_r2920_r2900_p756
	$t\bar{t}H$	116065	Pythia	PhotonFilter	e825_s1310_s1300_r2920_r2900_p756

Not unique values:
 could repeat in
 different projects

Campaign Description from AMI

```

...
}, {
  "MC11a": {
    "mc11_7TeV": {
      "digit": {
        "RDO": ["d579", "d580"]
      },
      "merge": {
        "AOD": ["r2700", "r2780"]
      },
      "recon": {
        "**": ["a128", "a131", "a133",
          "a134", ...]
      }
    }
  }
},
...
    
```

Project

mc11_7TeV.116866.***.e873_s1310_s1300_r2920_r2900_p756
 mc11_7TeV.125170.***.e893_s1310_s1300_r2920_r2900_p756
 mc11_7TeV.125329.***.e825_s1310_s1300_r2920_r2900_p756
 mc11_7TeV.125489.***.e825_s1310_s1300_r2920_r2900_p756
 mc11_7TeV.116064.***.e893_s1310_s1300_r2920_r2900_p756



Reconstruction of the real and MC datasets from document content

Measurement of underlying event characteristics using charged particle jets in pp collisions at $\sqrt{s} = 7\text{TeV}$ with the ATLAS detector at the LHC

All data used in this analysis were taken during the 2010 LHC running period A (run numbers 152166-153200) and period B (run numbers 153565-155160), with the May reprocessing (release 16).

Period's Description from AMI

```
[{"period": "A",
  "periodLevel": "2",
  "projectName": "data10_7TeV",
  "description": "unsqueezed stable beam data (beta*=10m): typical beam spot width in x and y is 50-60 microns.",
  "status": "locked"},
 {"period": "B",
  "periodLevel": "2",
  "projectName": "data10_7TeV",
  "description": "first squeezed stable beams (beta*=2m): typical beam spot width in x and y is 30-40 microns.",
  "status": "locked"}, {
```

Project Name
 "data10_7TeV"

Real data

generator	configuration	sample number
Pythia 6	AUET2B	<u>126169. 126346-126349</u>
Pythia 6	Z1	126172, 126358-126361
Pythia 6	Perugia2011	126170, 126354-126357
Pythia 6	Perugia2011 NOCR	126171, 126350-126353
Herwig++	UE7000	113906-113909
Pythia 8.145	4C	108316-108318, 108351, 113118-113125

Project Name
 "mc09_7TeV"

Table 2: EVGEN Monte Carlo samples

underlying physics distributions by using different MC *tunes* (configurations) as control samples. We use Pythia 6 (MC09) as the primary control sample because it has the best available statistics (19.6M events) compared to the data (42.6M events). More information about the MC samples can be found in Table 1.

MC





ATLAS NOTE

ATL-COM-PHYS-2010-685

March 9, 2011



Abstract

This paper presents the measurements of W , W^+ and W^- to muon and $Z \rightarrow \mu^+\mu^-$ inclusive cross-sections with the ATLAS detector in proton-proton collisions at $\sqrt{s}=7$ TeV. The results presented are based on an integrated luminosity of 310 nb^{-1} for the W analysis and 331 nb^{-1} for the Z analysis, collected in April-July 2010 with fully operational detector and stable beam conditions. There are 1181 W and 109 Z candidate events in the muon decay channel. The distributions for the main observables are compared to a PYTHIA Monte Carlo simulation at different stages of the selection. We measure $\sigma_W \times BR(W \rightarrow \mu\nu) = 9.58 \pm 0.30(\text{stat}) \pm 0.50(\text{sys}) \pm 1.05(\text{lum}) \text{ nb}$ and $\sigma_{Z/\gamma^*} \times BR(Z/\gamma^* \rightarrow \mu^+\mu^-) = 0.87 \pm 0.08(\text{stat}) \pm 0.05(\text{sys}) \pm 0.10(\text{lum}) \text{ nb}$, consistent with the Standard Model expectations.

$W \rightarrow \mu\nu$ and $Z \rightarrow \mu\mu$ cross-sections measurements in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS Detector

Process	Dataset	Generator	Cross-section (pb)	$N_{\text{evt}} (\times 10^6)$	note
$W \rightarrow \mu\nu$	106044	PYTHIA	10 454	7	
$Z \rightarrow \mu\mu$	106047	PYTHIA	989	7.9	$\sqrt{\hat{s}} > 60 \text{ GeV}$
$W \rightarrow \tau\nu$	106022	PYTHIA	10 454	5	single lepton filter ($\epsilon = 0.877$) times lepton branching ratio
$Z \rightarrow \tau\tau$	106052	PYTHIA	989	7.9	$\sqrt{\hat{s}} > 60 \text{ GeV}$
$i\bar{i}$	105861	POWHEG	161	0.2	$m_t = 172.5 \text{ GeV}/c^2$, single lepton filter $\epsilon = 0.538$
$b\bar{b}$	108405	PYTHIA	7.39×10^4	4.4	15 GeV/c single muon filter
$c\bar{c}$	106059	PYTHIA	2.84×10^4	1.5	15 GeV/c single muon filter
J0	105009	PYTHIA	9.75×10^9	0.4	$8 < \text{parton } p_T < 17 \text{ GeV}$
J1	105010	PYTHIA	6.73×10^8	0.4	$17 < \text{parton } p_T < 35 \text{ GeV}$
J2	105011	PYTHIA	4.12×10^7	0.4	$35 < \text{parton } p_T < 70 \text{ GeV}$
J3	105012	PYTHIA	2.19×10^6	0.4	$70 < \text{parton } p_T < 140 \text{ GeV}$
J4	105013	PYTHIA	8.79×10^4	0.4	$140 < \text{parton } p_T < 280 \text{ GeV}$
J5	105014	PYTHIA	2.33×10^3	0.4	$280 < \text{parton } p_T < 560 \text{ GeV}$
J6	105015	PYTHIA	3.39×10^2	0.4	$560 < \text{parton } p_T < 1120 \text{ GeV}$
J0mu	109276	PYTHIA	9.86×10^9	0.5	8 GeV/c single μ filter $\epsilon = 7.93 \times 10^{-5}$
J1mu	109277	PYTHIA	6.78×10^8	0.5	8 GeV/c single μ filter $\epsilon = 1.23 \times 10^{-3}$
J2mu	109278	PYTHIA	4.10×10^7	0.5	8 GeV/c single μ filter $\epsilon = 5.44 \times 10^{-3}$
J3mu	109279	PYTHIA	2.20×10^6	0.5	8 GeV/c single μ filter $\epsilon = 1.29 \times 10^{-2}$
J4mu	109280	PYTHIA	8.77×10^4	0.5	8 GeV/c single μ filter $\epsilon = 2.22 \times 10^{-2}$
J5mu	109281	PYTHIA	2.35×10^3	0.5	8 GeV/c single μ filter $\epsilon = 2.98 \times 10^{-2}$

Table 2: Monte Carlo samples used in this note. The cross-sections quoted are the ones used to normalize estimates of expected number of events. The cross-sections for the QCD samples ($b\bar{b}$, $c\bar{c}$, and the JX samples) are directly from PYTHIA. Sources for the other cross-sections are discussed in the text.

Run number range	Integrated Luminosity (nb^{-1})	
	W GRL	Z GRL
A-C: 152844-156682	16.65	17.60
D1: 158045-158392	26.89	28.64
D2: 158443-158582	29.03	31.76
D3: 158632-158975	32.85	34.71
D4: 158041-159086	79.40	87.82
D5: 159113	28.04	28.38
D6: 159179-159224	97.05	101.85
Total: 152844-159224	310.0	330.8

Table 1: Integrated luminosity for the runs in periods A to D for the W and Z Good Run Lists. The total integrated luminosity for this dataset is 310.0 nb^{-1} for the W and 330.8 nb^{-1} for the Z.



NodeJS Web Interface Prototype (previous)

Data Knowledge Base for ATLAS [Home](#) [About](#) [Contact](#)



Data Knowledge Base Prototype for ATLAS Collaboration. Processing papers & internal documents full text, search data samples, used in the data analysis.

Paper

document	http://nosql.tpu.ru:8890/DAV/ATLAS/document/95f92719-1075-418b-810b-ed2ab8c82223
type	http://nosql.tpu.ru/ontology/ATLAS#Paper
fullTitle	Search for a heavy narrow resonance decaying to e-mu, e-tau or mu-tau with the ATLAS detector in sqrt(s)=7 TeV pp collisions at the LHC
creationDate	2012-12-07T07:39:39
abstract	This letter presents the results of a search for a heavy particle decaying into an $e^+\mu^-$, $e^+\tau^-$, or $\mu^+\tau^-$ final state in pp collisions at $\sqrt{s} = 7$ TeV. The data were recorded with the ATLAS detector at the LHC during 2011 and correspond to an integrated luminosity of 4.5/fb. No significant excess above the Standard Model expectation is observed, and exclusions at 95% confidence level are placed on the cross section time branching ratio for the production of an R-parity-violating supersymmetric tau sneutrino. These results considerably extend constraints from Tevatron experiments.
DOI	10.1016/j.physletb.2013.04.035
GLANCE_ID	2645
ArXivCode	arXiv:1212.1272
CDSInternal	ATLAS-SUSY-2012-20-004
keywords_list	experimental results lepton production supersymmetry

Supporting Document

label	Supporting note
URL	https://cdsweb.cern.ch/record/1473263
GLANCE_ID	554
Get Datasets	

Datasets

- mc11_7TeV.105860.TTbar_PowHeg_Jimmy.digit.RDO.e873_s1310_s1300_d580_tid541913_00
- mc11_7TeV.105860.TTbar_PowHeg_Jimmy.digit.log.e873_s1310_s1300_d580_tid541913_00
- mc11_7TeV.105860.TTbar_PowHeg_Jimmy.evgen.EVNT.e1105_tid730643_00
- mc11_7TeV.105860.TTbar_PowHeg_Jimmy.evgen.EVNT.e1198_tid784995_00
- mc11_7TeV.105860.TTbar_PowHeg_Jimmy.evgen.EVNT.e1198_tid784996_00
- mc11_7TeV.105860.TTbar_PowHeg_Jimmy.evgen.EVNT.e1198_tid784995_00
- mc11_7TeV.105860.TTbar_PowHeg_Jimmy.evgen.EVNT.e873_tid491343_00
- mc11_7TeV.105860.TTbar_PowHeg_Jimmy.evgen.EVNT.e873_tid645914_00
- mc11_7TeV.105860.TTbar_PowHeg_Jimmy.evgen.EVNT.e873_tid648301_00
- mc11_7TeV.105860.TTbar_PowHeg_Jimmy.evgen.log.e1105_tid730643_00
- mc11_7TeV.105860.TTbar_PowHeg_Jimmy.evgen.log.e1198_tid784995_00
- mc11_7TeV.105860.TTbar_PowHeg_Jimmy.evgen.log.e1198_tid784996_00
- mc11_7TeV.105860.TTbar_PowHeg_Jimmy.evgen.log.e873_tid491343_00
- mc11_7TeV.105860.TTbar_PowHeg_Jimmy.evgen.log.e873_tid645914_00
- mc11_7TeV.105860.TTbar_PowHeg_Jimmy.evgen.log.e873_tid648301_00
- mc11_7TeV.105860.TTbar_PowHeg_Jimmy.merge.AOD.e1105_a131_s1353_a139_r2900_tid730658_00
- mc11_7TeV.105860.TTbar_PowHeg_Jimmy.merge.AOD.e1198_a131_s1353_a139_r2900_tid785122_00
- mc11_7TeV.105860.TTbar_PowHeg_Jimmy.merge.AOD.e1198_a131_s1353_a139_r2900_tid785123_00
- mc11_7TeV.105860.TTbar_PowHeg_Jimmy.merge.AOD.e1198_a131_s1353_a139_r2900_tid785124_00
- mc11_7TeV.105860.TTbar_PowHeg_Jimmy.merge.AOD.e1198_a131_s1353_a139_r2900_tid794416_00
- mc11_7TeV.105860.TTbar_PowHeg_Jimmy.merge.AOD.e873_a131_s1353_a139_r2900_tid600627_00
- mc11_7TeV.105860.TTbar_PowHeg_Jimmy.merge.AOD.e873_s1310_s1300_r2730_r2700_tid517950_00
- mc11_7TeV.105860.TTbar_PowHeg_Jimmy.merge.AOD.e873_s1310_s1300_r2730_r2700_tid517951_00
- mc11_7TeV.105860.TTbar_PowHeg_Jimmy.merge.AOD.e873_s1310_s1300_r2730_r2780_tid541655_00
- mc11_7TeV.105860.TTbar_PowHeg_Jimmy.merge.AOD.e873_s1310_s1300_r2730_r2780_tid541656_00
- mc11_7TeV.105860.TTbar_PowHeg_Jimmy.merge.AOD.e873_s1310_s1300_r3043_r2993_tid650480_00

