

EDISON Discussion Document

Common interchange format for information about courses and programs in Data Science

Project acronym: EDISON

Project full title: Education for Data Intensive Science to Open New science frontiers

Grant agreement no.: 675419

|  |  |
| --- | --- |
| Due Date | To be incorporated in D2.2 |
| Actual Date |  |
| Document Author/s | Tomasz Wiktorski |
| Version | 0.2 |
| Dissemination level | RESTRICTED (to be open when finalized) |
| Status | DRAFT |
| Document approved by |  |

|  |  |
| --- | --- |
|  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| Document Version Control | | | |
| Version | Date | Change Made (and if appropriate reason for change) | Initials of Commentator(s) or Author(s) |
| 0.1 | 18/02/2016 | Initial draft | TW |
| 0.2 | 3/02/2016 | Updated field lists and relation to existing standards | TW, AB, KB, HB, PW, RJ, YD |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

|  |  |  |
| --- | --- | --- |
| Document Change Commentator or Author | | |
| Author Initials | Name of Author | Institution |
| TW | Tomasz Wiktorski | University of Stavanger |
| KB | Kathrin Beck | Max Planck Computing and Data Facilty/RDA EU |
| AB | Adam Belloum | University of Amsterdam |
| HB | Holger Brocks | FTK |
| PW | Peter Wittenbug | Max Planck Computing and Data Facilty/RDA EU |
| RJ | Rafael Jimenez | Elixir |
| YD | Yuri Demchenko | University of Amsterdam |

|  |  |
| --- | --- |
|  |  |

# Motivation and instructions

This goal of this format is to simplify publishing and gathering of information about courses and programs in data science and related domains. It is purposefully very generic to accommodate for a wide variety of courses, both regular and one-off.

We included the field that we consider to be important both for general informational purposes but also from perspective of education theory and alignment with EDISON’s Data Science Competence Framework (CF-DS), Body of Knowledge (CF-BoK), RDA EU’s Training Specification, CODATA?, Elixir?. In case the content for some of the fields might be difficult to obtain for all courses and programs, we suggest keeping them in the format but making them optional.

On the next two pages you can find list of suggested field for Courses (subjects) and Programs. This are combined from original version and comments received. I included all the suggestions, but sometimes name of the field might be different than you requested.

This is a review document; please evaluate it from the perspective of your organization. Edit in the review mode and append name of your organization to file name.

Please consider the following questions when reviewing the document:

1. Is any field missing? If so, could you use one of the existing field for it?
2. Do you agree with Obligatory, Recommended, Optional recommendations?
3. What is your opinion on existing standards? Do you use them? If no, why not?

# Courses (Subjects)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Field Name | Obligatory  Recommended  Optional | iCal | Schema.org/event | XCRI-CAP | LRMI | Description |
| Title | Obligatory | + | + | + | + | A meaningful short title |
| Name of Presenter(s) | Optional | - | + | - | + | A person of a list of people delivering the course, with their affiliations |
| Organizer | Obligatory | - | + | + | +/- | Institution, company, project organizing the course |
| Type of Course | Optional | - | - | + | + | Webinar, summer school, academic course |
| Location | Recommended | + | - | + | - | A country and city where course takes place, unless online |
| Start Date and Time | Obligatory | + | + | + | - | The start date and time of the item (in ISO 8601 date format). |
| End Date and Time | Obligatory | + | + | + | - | The end date and time of the item (in ISO 8601 date format). |
| URL | Obligatory | + | + | + | + | Link to further information |
| Contact | Obligatory | + | +/- | +/- | - | A person/email that should be used for contacting |
| Language | Obligatory | - | + | + | + |  |
| Level | Optional | - | - | +/- | - | Which level of studies following either Bologna or US approach |
| Credit | Recommended | - | - | + | - | Recommended for academic courses |
| Location | Optional | + | + | + | - | Institution, address, name of host organization |
| Prerequisites | Recommended | - | - | + | +/- | Required prior knowledge, preferably based on a BoK or taxonomy |
| Target Audience | Optional | - | - | - | - |  |
| Learning Outcomes | Recommended | - | - | + | - | Including objectives, preferably based on a CF |
| Description | Optional | + | + | + | + | The course will provide a strong basis in administrative, programing, and algorithm design aspects of data intensive systems. |
| Registration Deadline | Optional | - | - | +/- | - | The date and time of the item (in ISO 8601 date format). |
| Cost | Optional | - | - | + | - | Use three letter currency symbols |

# Programs

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Field Name | Obligatory  Recommended  Optional | iCal | Schema.org/event | XCRI-CAP | LRMI | Description |
| Title | Obligatory | + | + | + | + | A meaningful short title |
| Track Name | Optional | - | - | - | - | Name of the track within the program |
| Course List | Recommended | - | - | +/- | +/- | URI to courses being part of the program, limited to the track if specified |
| Organizer | Obligatory | - | + | + | +/- | Institution, company, project organizing the course |
| Type of Course | Optional | - | - | + | + | Webinar, summer school, academic course |
| Location | Recommended | + | - | + | - | A country and city where course takes place, unless online |
| Start Date and Time | Obligatory | + | + | + | - | The start date and time of the item (in ISO 8601 date format). |
| End Date and Time | Obligatory | + | + | + | - | The end date and time of the item (in ISO 8601 date format). |
| URL | Obligatory | + | + | + | + | Link to further information |
| Contact | Obligatory | + | +/- | +/- | - | A person/email that should be used for contacting |
| Language | Obligatory | - | + | + | + |  |
| Level | Optional | - | - | +/- | - | Which level of studies following either Bologna or US approach |
| Credit | Recommended | - | - | + | - | Recommended for academic courses |
| Location | Optional | + | + | + | - | Institution, address, name of host organization |
| Prerequisites | Recommended | - | - | + | +/- | Required prior knowledge, preferably based on a BoK or taxonomy |
| Target Audience | Optional | - | - | - | - |  |
| Learning Outcomes | Recommended | - | - | + | +/- | Including objectives, preferably based on a CF |
| Description | Optional | + | + | + | + | The course will provide a strong basis in administrative, programing, and algorithm design aspects of data intensive systems. |
| Registration Deadline | Optional | - | - | +/- | - | The date and time of the item (in ISO 8601 date format). |
| Cost | Optional | - | - | + | - | Use three letter currency symbols |

# Comments on standards

Review of the existing standards demonstrates that no single standard would fully cover all needs specified by project partners. In some cases, there is no explicit coverage for a particular field, but there are some closely related fields, I tried to reflect these in the tables.

It is important to represent relevant information with iCalendar standard. Such approach would facilitate easy import into various calendar applications people might be using.

Schema.org/event does not offer much coverage over iCalendar. Considering iCalendar adoption in calendar applications, schema.org/event does not seem to be useful for our purpose.

XCRI-CAP covers the majority of requested fields. However, after initial review, it seems to be a fairly complicated standard. Despite most fields are covered; this coverage is often indirect or required additional structures and information which do not seem to be necessary for our purpose. The complexity of XCRI-CAP might be a hurdle in adoption, especially for our purpose.

LRMI does not cover as much fields as XCRI-CAP; however, it seems to be more straight-forward to use.

My suggestion for discussion would be to extend LRMI and provide some form of integration with iCalendar simultaneously. As for the semantic specification of such approach, an expert opinion should be sought.

APARSEN and FOSTER needs were also reviewed based on documents provided. It appears that current proposal should cover their needs. More input from project participants is appreciated.

# Appendixes

These are old XML example for refence only. Do not reflect newest fields.

**Course Example in XML**

<?xml version="1.0" encoding="UTF-8"?>

<dscourse courseid="" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLocation="dscourse.xsd">

<name>Data Intensive Systems</name>

<institution>Univeristy of Stavanger</institution>

<country>Norway</country>

<unit>Department of Electrical and Computer Engineering</unit>

<link>http://www.uis.no/studies/study-courses/?categoryID=10648&parentcat=9835&code=DAT500\_1&name=Data-Intensive+systems</link>

<contact>Tomasz Wiktorski (tomasz.wiktorski@uis.no)</contact>

<language>English</language>

<level>Graduate/Master</level>

<credit>10 ECTS</credit>

<date> </date>

<place>University of Stavanger</place>

<description>The course will provide a strong basis in administrative, programing, and algorithm design aspects of data intensive systems.</description>

<learningoutcomes>Characterize Hadoop job tracker, task tracker, scheduling issues, communications, and resource management. Describe elements of Hadoop ecosystem and identify their applicability. Describe and compare RDBMS, data warehouse, unstructured big data, and keyed files, and show how to apply them to typical data processing problems. Understand algorithmic complexity of the worst case, expected case, and best case running time, and the orders of complexity; apply the analysis to real life algorithms. Design, construct, test, and benchmark a small data processing cluster (based on Hadoop). Analyze real-life problems and propose suitable solutions. Construct programs based directly on MapReduce paradigm for typical problems. Construct programs based on high-level tools (for MapReduce paradigm) for typical problems. Analyze influence of peak and sustained bandwidth rate on system performance. Evaluate, communicate and defend a data-intensive solution w.r.t. relevant criteria.</learningoutcomes>

</dscourse>

**Course XML schema for field definition**

dscourse.xsd

<?xml version="1.0" encoding="UTF-8" ?>

<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">

<xs:element name="dscourse">

<xs:complexType>

<xs:attribute name="courseid" type="xs:string" use="required"/>

<xs:sequence>

<xs:element name="name" type="xs:string"/>

<xs:element name="institution" type="xs:string"/>

<xs:element name="country" type="xs:string"/>

<xs:element name="unit" type="xs:string" minOccurs="0"/>

<xs:element name="link" type="xs:string"/>

<xs:element name="contact" type="xs:string"/>

<xs:element name="language" type="xs:string"/>

<xs:element name="level" type="xs:string" minOccurs="0"/>

<xs:element name="credit" type="xs:string" minOccurs="0"/>

<xs:element name="date" type="xs:string" minOccurs="0"/>

<xs:element name="place" type="xs:string" minOccurs="0"/>

<xs:element name="description" type="xs:string" minOccurs="0"/>

<xs:element name="learningoutcomes" type="xs:string" minOccurs="0"/>

</xs:sequence>

</xs:complexType>

</xs:element>

</xs:schema>

**Program Example in XML**

<?xml version="1.0" encoding="UTF-8"?>

<dsprogram programid="" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLocation="dsprogram.xsd">

<name>Data Science</name>

<institution>Danish Technical University</institution>

<country>Denmark</country>

<unit>Department of Applied Mathematics and Computer Science</unit>

<link>http://www.compute.dtu.dk/english/education/Data-Science-Big-Data</link>

<contact>Bjarne Kjær Ersbøll (bker@dtu.dk)</contact>

<language>English</language>

<level>Graduate/Master</level>

<title>MSc.</title>

<description>The obtained profile provides thorough competencies in the analysis of massive data, which cover the whole spectrum from data acquisition through storage, analysis and interpretation to the application and presentation of the results.</description>

<learningoutcomes></learningoutcomes>

</dsprogram>

**Program XML Schema for field definition**

<?xml version="1.0" encoding="UTF-8" ?>

<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">

<xs:element name="dsprogram">

<xs:complexType>

<xs:attribute name="programid" type="xs:string" use="required"/>

<xs:sequence>

<xs:element name="name" type="xs:string"/>

<xs:element name="institution" type="xs:string"/>

<xs:element name="country" type="xs:string"/>

<xs:element name="unit" type="xs:string" minOccurs="0"/>

<xs:element name="link" type="xs:string"/>

<xs:element name="contact" type="xs:string"/>

<xs:element name="language" type="xs:string"/>

<xs:element name="level" type="xs:string" minOccurs="0"/>

<xs:element name="title" type="xs:string" minOccurs="0"/>

<xs:element name="description" type="xs:string" minOccurs="0"/>

<xs:element name="learningoutcomes" type="xs:string" minOccurs="0"/>

</xs:sequence>

</xs:complexType>

</xs:element>

</xs:schema>