MyData Open Data Specification

Version 1.3
May 22, 2012

Overview

As a society that relies increasingly on electronic systems to gather and deliver data in nearly all of our activities, providing consumer-generated data back to the consumer expands product and service choices, enables better informed decisions and optimizes the marketplace. The Obama Administration is promoting opening access of usage data back to the consumer (termed "smart disclosure") as a core component of its Open Government Directive. The principals behind smart disclosure are summarized in the September 8, 2011 memorandum titled "Informing Consumers through Smart Disclosure" [PDF].

The call for smart disclosure has been led by the U.S. Department of Veterans Affairs (VA) Blue Button and US Department of Energy (DOE) Green Button initiatives. The Blue Button initiative allows for veterans to view and download their personal health records in a plain text format. A number of companies are offering derivative solutions that utilize this data, including viewing data on mobile devices and applying for jobs using military service data. The Green Button initiative allows consumers to download their home energy usage data from their electricity utility provider. This initiative was recently launched in March 2012 and their upcoming $100,000 developer contest is expected to bring many innovative applications to deliver enriched value to consumers from their usage data.

MyData Button

Inline with the Blue and Green Button initiatives, The US Department of Education is facilitating an industry movement to connect education consumers with their data called the MyData Button. This initiative calls educational entities (schools, universities, test preparation companies and etc) and service providers to offer an easy-to-use data download button for the customers they service. With this data available to the consumer, we expect the educational technology industry to create a new wave of useful applications to open new market opportunities for college completion, financial planning, customized learning and classroom preparation for teachers. The goal of this initiative is to a.) create a simple to use mechanism for consumers to download their educational data in a machine readable format on a provider's web site and b.) define a minimal open data format for a provider to offer this functionality with minimal implementation burden by leveraging data standards they already in use.

Format Proposal

Educational providers should be able to choose any machine-readable standard suitable to the data which they are representing. A number of open education formats exist including the Schools Interoperability Framework (SIF), Postsecondary Electronic Standards Council (PESC) and IMS Global Learning Consortium (IMS) to name a few. The recommended format to encapsulate these documents into a single MIME message using the multipart media type as defined by RFC 2045 and RFC 2046 Section 5.1 "The Multipart Media Type". This specification has been used since 1996 in SMTP email and NNTP newsgroups and defines a very simple container equally usable for unstructured plain text data, structured data (XML, JSON, etc) or binary data (PDF, JPG, etc) in either well-known or arbitrary formats. A multipart media type used by the MyData Button initiative will offer the following benefits:

- **Ease of Implementation** - To define data within a multipart media type, one plain text line is required to declare the data type. Directly below the content type declaration is the data itself, which can be plain text, XML or binary data (base 64 encoded).

- **Unlimited Number of Attachments** - Multipart content types support an unlimited number of attachments, only constrained by disk space and reasonable data transfer practices. In a single data download, it would be possible to enclose a number of plain text documents; structured XML records; and binary attachments (image, PDF and Word documents) representing a full student profile.

- **Flexible and Extensible Format Support** - MIME multipart content types can accommodate any text or binary attachment type. The format allows for current standards today (SIF, PESC, IMS, as well as images and documents) and will accommodate any new standards in the future. For those who have data but not in a standardized format, generic types are available to represent the data.

Defining a Multipart Content Type

Within a MIME multipart media type, formatted plain lines are used to denote the contents within.

The start Content-Type line, indicating that the file will contain multiple parts of mixed data types. The boundary field denotes that "--education_mydata" will be used as a separator between attached data sets:

```
Content-Type: multipart/mixed; boundary=education_mydata
```

Below is a data separator as defined by the boundary field above. A data separator is inserted before each data set attachment. It must begin on a new line and must be preceded by a blank line:

```
--education_mydata
```

Each data attachment is prefixed with a line Content-Type defining its media type. The Content-Type line is followed by a blank line and then the content.

Media types are registered by the Internet Assigned Numbers Authority. The Content type header is defined by RFC 2045 which states "The Content-Type field is used to specify the nature of the data in the body of a MIME entity, by giving media type and subtype identifiers, and by providing auxiliary information that may be required for certain media types." Top-level defined media types are defined by RFC 2046. A listing of registered MIME media types is available here: http://www.iana.org/assignments/media-types/index.html.

Media types can be generic such as "text/plain" for plain text, "text/xml" for XML data or "application/octet-stream" for binary data. Media types can be specific to a known file type such as "application/pdf" for Adobe PDF or "application/msword" for Microsoft Word. For application specific types, please see the "Defining MIME Types" section below.

Examples of MIME media types in format as specified by RFC 2045:

```
Content-Type: text/plain
Content-Type: text/xml
Content-Type: application/vnd.sif.academicrecord+xml
Content-Type: application/vnd.pesc.transcript+xml
Content-Type: application/pdf
```

Declaring MIME Content Types

The Internet Assigned Numbers Authority (IANA) is the body that oversees media types. An organization desiring to register a media type, an [Application for Media Type](#) must be filed. New media types SHALL be of type "application", subtype vendor ("vnc"), vendor designation (i.e. "sif"), name of the document type (i.e. "academicrecord") and its media type (i.e. "xml"). The media type SHALL be within be within the "vnd" tree consisting of the vendor name and the document type. For example, `application/vnd.sif.academicrecord+xml` would represent a SIF academic record in XML.

IANA registration SHOULD occur for all vendor document formats, however IANA registration is NOT REQUIRED to utilize unused vendor and document types. When determining media type naming conventions for an organization, review the IANA media type registration list at http://www.iana.org/assignments/media-types/application/index.html to ensure the convention has not been previously registered. Unused vendor and document types can immediately be used for denoting the contents of a MyData download.

Please note, document version information is contained within the document itself, not in the media type.

Delivering MyData

To ensure the end-user has a consistent and secure experience when downloading from the MyData initiative, the following specifications provide guidance for delivering data:

- All data interaction MUST occur over Hypertext Transfer Protocol 1.1 (HTTP) using Transport Layer Security (TLS) as a cryptographic protocol. Data interaction includes any instance a user may submit or receive sensitive or personally identifiable information such as an account number, password, social security number, academic history and so on. Transferring over HTTP using TLS will ensure the communication channel is encrypted between the information provider and the end user, protecting the end user from eavesdropping and tampering by unintended parties. HTTP and TLS are common to all major web browsers and servers in the marketplace, so should not pose a burden to the implementor. An implementor can place a HTTPS URI linking to the data on a web site. When the user clicks on the URI link, the web server will use HTTP and TLS to transfer the data to the user.

- When delivering a MyData file to the end-user (i.e. the end-user downloads a MyData file) the following two HTTP header parameters MUST be set. This will notify the consuming application of file type as well as force the file to download instead of displaying as text in the web browser.

    - a.) the HTTP Content-Type header from the web server MUST be set to "multipart/mixed" and MUST define a boundary (such as "education_mydata" as shown below) ([defined by IETF RFC 2045](#)). The boundary name is arbitrarily determined by the implementor as along as consistent within the document.

    - b.) the HTTP Content-Disposition header MUST be set to "attachment" and a filename SHOULD be set ([defined by IETF RFC 2183](#)).

    As a result, the additional HTTP headers for a MyData file will read:

    ```
    Content-Type: multipart/mixed; boundary=education_mydata;
    Content-Disposition: attachment; filename=mydata.dat;
    ```

    Samples in popular web programming languages are provided below:

    - **ASP.NET:**

```
Response.AddHeader("Content-Type", "multipart/mixed; boundary=education_mydata");
Response.AddHeader("Content-Disposition", "attachment; filename=mydata.dat");
```

○ **JSP:**

```
response.setHeader("Content-Type", "multipart/mixed; boundary=education_mydata");
response.setHeader("Content-Disposition", "attachment; filename=mydata.dat");
```

○ **PHP:**

```
header('Content-Type: multipart/mixed; boundary=education_mydata');
header('Content-Disposition: attachment; filename=mydata.dat');
```

● It is RECOMMENDED that MyData files are sent by the web server using GZIP compression. Given the MyData format is text (optionally with text encoded binaries), this will reduce download time and data transferred by up to 70% offering the end-user a faster download and saving bandwidth costs for implementors.

Example

Below is an example of sample student record (the entire HTTP response is not shown, only the message body) containing:

a.) Plain text file of a college attendance record.
b.) SIF XML student academic record.
c.) PESC XML college transcript.
d.) Adobe PDF binary attachment of a student record.

```
Content-Type: multipart/mixed; boundary=education_mydata

--education_mydata
Content-type: text/plain; charset=us-ascii

NAME:                   Jane Doe
STREET1:        123 Main Street
CITY:                   Anywhere
STATE:          CA

COLLEGE:        California State University
START_YEAR:     2002
END_YEAR:       2006
DEGREE_LEVEL:Master, Arts
DEGREE:                 Education
GPA:            3.85

--education_mydata
Content-type: application/vnd.sif.academicrecord+xml

<?xml version="1.0" encoding="UTF-8"?>
<StudentAcademicRecord RefId="D932E34082D04C048112729922B58A93"
SIF_RefObject="StudentRecordExchange" SIF_RefId="0EAF9D6DA4854F348103B04BFEAC65B2">
  <ReportingDate>2006-09-26</ReportingDate>
  <StudentSchoolEnrollmentData SchoolAttendedRefId="39FA2714A63543B0A7732796825E9C68">
    <GradeLevel>
      <Code>12</Code>
    </GradeLevel>
  </StudentSchoolEnrollmentData>
```

```xml
<DistrictEntryDate>2003-09-01</DistrictEntryDate>
<SchoolAttendanceHistory>
  <SchoolAttended RefId="39FA2714A63543B0A7732796825E9C67">
    <SchoolInfoData>
      <LocalId>43</LocalId>
      <StateProvinceId>389</StateProvinceId>
      <SchoolName>Thomas Jefferson HS</SchoolName>
      <GradeLevels>
        <GradeLevel>
          <Code>09</Code>
        </GradeLevel>
        <GradeLevel>
          <Code>10</Code>
        </GradeLevel>
        <GradeLevel>
          <Code>11</Code>
        </GradeLevel>
        <GradeLevel>
          <Code>12</Code>
        </GradeLevel>
      </GradeLevels>
    </SchoolInfoData>
    <MarkingSystems>
      <MarkValueInfoData RefId="DBA34A13FF234C7994AAE13D89DA9A50">
        <Name>Letter Grades</Name>
        <Letter>
          <ValidMark>
            <Code>A</Code>
            <NumericEquivalent>4.0</NumericEquivalent>
          </ValidMark>
          <ValidMark>
            <Code>A-</Code>
            <NumericEquivalent>3.7</NumericEquivalent>
          </ValidMark>
          <ValidMark>
            <Code>B+</Code>
            <NumericEquivalent>3.2</NumericEquivalent>
          </ValidMark>
          <ValidMark>
            <Code>B</Code>
            <NumericEquivalent>3.0</NumericEquivalent>
          </ValidMark>
          <ValidMark>
            <Code>B-</Code>
            <NumericEquivalent>2.7</NumericEquivalent>
          </ValidMark>
          <ValidMark>
            <Code>C+</Code>
            <NumericEquivalent>2.8</NumericEquivalent>
          </ValidMark>
          <ValidMark>
            <Code>C</Code>
            <NumericEquivalent>2.0</NumericEquivalent>
          </ValidMark>
          <ValidMark>
            <Code>C-</Code>
            <NumericEquivalent>1.7</NumericEquivalent>
          </ValidMark>
          <ValidMark>
            <Code>D+</Code>
            <NumericEquivalent>1.2</NumericEquivalent>
```

```xml
        </ValidMark>
        <ValidMark>
          <Code>D</Code>
          <NumericEquivalent>1.0</NumericEquivalent>
        </ValidMark>
        <ValidMark>
          <Code>E</Code>
          <NumericEquivalent>0.0</NumericEquivalent>
        </ValidMark>
        <ValidMark>
          <Code>F</Code>
          <NumericEquivalent>0.0</NumericEquivalent>
        </ValidMark>
      </Letter>
    </MarkValueInfoData>
  </MarkingSystems>
</SchoolAttended>
</SchoolAttendanceHistory>
</StudentAcademicRecord>


--education_mydata
Content-type: application/vnd.pesc.transcript+xml

<?xml version="1.0" encoding="UTF-8"?>
<ColTrn:CollegeTranscript
      xmlns:ColTrn="urn:org:pesc:message:CollegeTranscript:v1.2.0"
      xmlns:AcRec="urn:org:pesc:sector:AcademicRecord:v1.5.0"
xmlns:core="urn:org:pesc:core:CoreMain:v1.8.0"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      xsi:schemaLocation="urn:org:pesc:message:CollegeTranscript:v1.2.0
CollegeTranscript_v1.2.0.xsd">
      <TransmissionData>
            <DocumentID>050330001</DocumentID>
            <CreatedDateTime>2005-02-02T10:04:29-05:00</CreatedDateTime>
            <DocumentTypeCode>RequestedRecord</DocumentTypeCode>
            <TransmissionType>Original</TransmissionType>
            <Source>
                  <Organization>
                        <ATP>006882</ATP>
                        <LocalOrganizationID>
                              <LocalOrganizationIDCode>123456</LocalOrganizationIDCode>
                              <LocalOrganizationIDQualifier>TX</LocalOrganizationIDQualifier>
                        </LocalOrganizationID>
                        <OrganizationName>THE UNIVERSITY OF TEXAS AT AUSTIN
                        </OrganizationName>
                        <Contacts>
                              <Address>
                                    <AddressLine>Office of the Registrar</AddressLine>
                                    <City>Austin</City>
                                    <StateProvinceCode>TX</StateProvinceCode>
                                    <PostalCode>78712</PostalCode>
                                    <AttentionLine>Jane Shaugness</AttentionLine>
                              </Address>
                              <Phone>
                                    <AreaCityCode>512</AreaCityCode>
                                    <PhoneNumber>4757638</PhoneNumber>
                              </Phone>
                              <FaxPhone>
                                    <AreaCityCode>512</AreaCityCode>
                                    <PhoneNumber>4757600</PhoneNumber>
```

```xml
                                </FaxPhone>
                                <Email>
                                        <EmailAddress>Registrar@mail.utexas.edu</EmailAddress>
                                </Email>
                                <URL>
                                        <URLAddress>www.utexas.edu</URLAddress>
                                </URL>
                        </Contacts>
                        <Accreditation>
                                <AccreditingBodyCode>SACS</AccreditingBodyCode>
                                <AccreditingBodyName>Southern Association of Colleges and
                                        Schools
                                </AccreditingBodyName>
                        </Accreditation>
                </Organization>
        </Source>
        <Destination>
                <Organization>
                        <FICE>001506</FICE>
                        <LocalOrganizationID>
                                <LocalOrganizationIDCode>1234</LocalOrganizationIDCode>
                                <LocalOrganizationIDQualifier>FL</LocalOrganizationIDQualifier>
                        </LocalOrganizationID>
                        <OrganizationName>Miami Dade College</OrganizationName>
                </Organization>
        </Destination>
        <RequestTrackingID>23456778009</RequestTrackingID>
</TransmissionData>
<Student>
        <Person>
                <SSN>100000000</SSN>
                <Birth>
                        <BirthDate>1967-07-14</BirthDate>
                </Birth>
                <Name>
                        <NameCode>Legal</NameCode>
                        <FirstName>Jane</FirstName>
                        <MiddleName>Kay</MiddleName>
                        <LastName>Doe</LastName>
                </Name>
                <HighSchool>
                        <OrganizationName>AUSTIN HIGH SCHOOL</OrganizationName>
                        <CEEBACT>440320</CEEBACT>
                </HighSchool>
                <Contacts>
                        <Address>
                                <AddressLine>1223 University Dr</AddressLine>
                                <City>Austin</City>
                                <StateProvinceCode>TX</StateProvinceCode>
                                <PostalCode>78211</PostalCode>
                                <ExpirationDate>2009-06-15</ExpirationDate>
                        </Address>
                        <Email>
                                <EmailAddress>TestSpeede@aol.com</EmailAddress>
                        </Email>
                </Contacts>
                <Gender>
                        <GenderCode>Female</GenderCode>
                </Gender>
        </Person>
        <AcademicRecord>
```

```
                <School>
                        <OrganizationName>AUSTIN HIGH SCHOOL</OrganizationName>
                        <CEEBACT>440320</CEEBACT>
                        <SchoolLevel>Secondary</SchoolLevel>
                        <NoteMessage>LOCATION: AUSTIN, TX</NoteMessage>
                </School>
                <AcademicAward>
                        <AcademicAwardLevel>B18</AcademicAwardLevel>
                        <AcademicAwardDate>1985-05-01</AcademicAwardDate>
                        <AcademicAwardTitle>Standard High School Diploma
                        </AcademicAwardTitle>
                        <AcademicCompletionIndicator>true</AcademicCompletionIndicator>
                </AcademicAward>
            </AcademicRecord>
        </Student>
</ColTrn:CollegeTranscript>

--education_mydata
Content-type: application/pdf
```

```
VGhlIHB1cnBvc2Ugb2YgIHRoaXMgTWVtb3JhbmR1bSBpcyB0byBzZXQgb3V0IGd1aWRhbmNlIGZv
ciBhZ2VuY2llcyB0byBpbmZvcm0gYW5kIA0KZmFjaWxpdGF0ZSB0aGUgdXNlIG9mICBkaXNjbG9z
dXJlLCBzcGVjaWZpY2FsbHkgInNtYXJ0IGRpc2Nsb3N1cmUuIiBBcyB1c2VkIGhlcmUsIHRoZSB0
ZXJtICJzbWFydCAgDQpkaXNjbG9zdXJlIiByZWZlcnMgdG8gdGhlIHRpbWVseSByZWxlYXNlIG9m
ICBjb21wbGV4IGluZm9ybWF0aW9uIGFuZCBkYXRhIGluIHN0YW5kYXJkaXplCwgbWFjaGluZSAN
CnJlYWRhYmxlIGZvcm1hdHMgaW4gd2F5cyB0aGF0IGVuYWJsZSBjb25zdW1lcnMgdG8gbWFrZSBp
bmZvcm1lZCBkZWNpc2lvbnMuIFNtYXJ0IGRpc2Nsb3N1cmUgDQp3aWxsIHR5cGljYWxseSB0YWtl
IHRoZSBmb3JtIG9mICBwcm9haWRpbmcgaW5kaXZpZHVhbCBjb25zdW1lcnMgb2YgIGdvb2RzIGFu
ZCBzZXJ2aWNlcyB3aXRoIGRpcmVjdCANCmFjY2VzcyB0byByZWxldmFudCBpbmZvcm1hdGlvbiBh
bmQgZGF0YSBzZXRzLiBTdWNoIGluZm9ybWF0aW9uIG1pZ2h0IGludm9sdmUsIGZvciBleGFtcGxl
LCB0aGUgDQpyYW5nZSBvZiAgY29zdHMgYXNzb2NpYXRlZCB3aXRoIHZhcmlvdXMgcHJvZHVjdCAN
CmFuZCHNlcnZpY2VzLCBpbnNddWRpbmcgY29zdHMgdGhhdCBtaWdodCBub3QgDQpvdGhlcndpc2Ug
YmUgdHJhbnNwYXJlbnQuIEluIHNvbWUgY2FzZXMsIGFnZW5jaWVzIG9yIHRoaXJkLXBhcnR5IGlu
dGVybWVkaWFyaWVzIG1heSBhbHNvIGNyZWF0ZSANCnRvb2xzIHRoYXQgdXNlIHRoZXNlIGRhdGEg
c2V0cyB0byBwcm9haWRlIHNlcnZpY2VzIHRoYXQgc3VwcG9ydCBjb25zdW1lciBkZWNpc2lvbi1t
YWtpbmcuIFN1Y2ggDQpkZWNpc2lvbi1tYWtpbmcgbWlnaHQgYmUgIGltcHJvdmVkLCBmb3IgZXhh
bXBsZSwgYnkgaW5mb3JtaW5nIGNvbnN1bWVyeyBhYm91dCB0aGUgbmF0dXJlIGFuCANCmVmZmVj
dHMgb2YgIHRoZWlyIG93biBwYXN0IGRlY2lzaW9ucyAoaW5jbHVkaW5nLCBmb3IgZXhhbXBsZSwg
dGhlIGNvc3RzIGFuZCBmZWVzIHRoZXkgaGF2ZSBhbHJlYWR5IA0KaW5jdXJyZWQpLg==
```

Change Log
- April 7, 2012 - Version 0.5 - Initial Release and Request for Comments
- May 8, 2012 - Version 1.0 - Added language and example for HTTP Content-Type header and GZIP compression
- May 11, 2012 - Version 1.1 - Modifications and comments for technical clarity
- May 16, 2012 - Version 1.2 - Added language and example for HTTP Content-Disposition header and modifications for technical clarity
- May 22, 2012 - Version 1.3 - Clarifications in MIME type usage and registrations