EMREX

Recommendations for developers

Janina Mincer-Daszkiewicz
2017-12-30

EMREX project is co-funded by the Erasmus+ Programme of the European Union under the grant 388499-EPP-1-2014-2-FI-EPPKA3-PI-POLICY. The University of Warsaw is also co-financed by the Polish Ministry of Science and Higher Education from the funds allocated in the years 2016-2017 for science, granted to international co-funded project.
Tabel of Contents

1. Introduction .................................................................................................................................................. 2
2. Recommendations for developers .................................................................................................................. 3
   2.1. Software development ............................................................................................................................... 3
   2.2. Acceptance testing .................................................................................................................................... 4
   2.3. Problems to solve, best practices, lessons learnt ...................................................................................... 4
   2.4. Automatic recognition .............................................................................................................................. 5
   2.5. ELMO backward compatibility .................................................................................................................. 6
3. Anticipation of potential risks ........................................................................................................................ 7

1. Introduction

**EMREX** is a proven solution for electronic transfer of student records between higher education institutions in Europe. Its objective is to increase availability, quality and reliability of information about student records of achievement and thus to make the administration of student mobility easier.

The EMREX Network consists of the following elements (see [http://emrex.eu/technical/](http://emrex.eu/technical/)):

1. **EMREX Client** – this is the web application used by a student to initiate the transfer of the achievements from another country.
2. **SMP (Student Mobility Plugin)** – this is a plugin that the EMREX client uses to enable the communication with an NCP and to ensure that it is done in a standardized way.
3. **NCP (National Contact Point)** – this is the point that the EMREX client contacts to fetch the student’s results. It is responsible for collecting and sending the data. Despite the name, one NCP may support one institution, a couple of institutions or all institutions in the country.
4. **EMREG** – the EMREX Registry. It is the only central component of the network. It keeps the list of all available NCPs as well as other information necessary to establish secure communication with each of them (e.g. certificates). Upon request it returns a list of all available countries and NCPs, where each NCP can contain a list of all institutions it is responsible for.

To join the Network (see [http://emrex.eu/steps-to-join/](http://emrex.eu/steps-to-join/)) the institution should develop:

1. Own SMP and the Client – to get the data from the partner institutions. The EMREX User Group offers a common implementation of SMP to help countries developing the Clients. This SMP can be used in existing web applications that need results from other HEIs, like applications for mobility students or admission systems.
Recommended for developers

2. Own NCP – to provide the data to the partner institutions. Each country participating in the EMREX Network must implement one or more NCPs, depending on the existing infrastructure in the specific country.

The EMREX Network is decentralised. Some software elements are open source, examples for implementations are available for different program languages and frameworks. Industry standard has been implemented for the data exchange and security. Integrity and confidentiality is essential for this kind of network and taking care of this is an important part of the solution. It has also been important to come up with a solution that is relatively easy to connect to from each HEI, and therefore support the upscale of the EMREX Network.

This document contains some recommendations for developers of local solutions needed by higher education institutions or other organizations, like e.g. admission agencies, planning to join the EMREX Network.

2. Recommendations for developers

2.1. Software development

It is recommended that the development team first gets to know the Technical guide and in particular the EMREX for Dummies document. Open source parts of the EMREX solution are stored in GitHub.

During software requirements analysis issues concerning usability of the developed software and user experience should be taken into account. Another important feature is scalability (from technical and financial perspective). It should be easy to set up an NCP in other countries using the same or similar software. Contact Points could be built using open source technologies, so that their deployment does not involve extra license costs. Data privacy and security is crucial, as the data is sent across borders. Each different SIS needs a separate SMP, but some general architectural decisions may be reused among various solutions. Technical issues should be discussed with the partners and the solution allowing reuse of software should be selected, if possible.

Software should be developed according to software requirements specification. It should be tested if functional and non-functional requirements have been met. These are the most important functionalities and features:

- Users can be authorized and identified with no special effort. In particular students are recognized as data owners and can trigger the action of data transfer for recognition purposes.
- Data concerning student achievements can be sent from one institution to the other (usually from a receiving institution to a sending institution).
- Privacy of data is protected.
- System is user-friendly, i.e. occasional users do not need special trainings to use it effectively and system responses to user requests in a reasonable time.
- System is robust and scalable, e.g. can be easily deployed in other institutions.
The EMREX solution is based on the ELMO standard. Mapping data from the local database to the ELMO format is one of the most challenging parts of the development. It should be remembered that even if some fields are not obligatory, they may be expected by the partners. Preparing the ELMO XML file from the local data we should deliver as much data as possible. Created XML should be validated against the EMREX ELMO Schema. The Online Debugger/Validator can be used for that purpose. The XML file can be fetched from the existing NCP server or uploaded manually.

The detailed description of that EMREX implementation in Poland is available in the documentation. EMREX Client and NCP have been integrated with USOSweb which is student’s portal used in Polish Higher Education Institutions gathered in the consortium MUCI. USOSweb is part of the family of applications which constitute Student Information System called USOS. The document describes various scenarios with screen shots which may be helpful when designing own solution. All the scenarios where taken in the development environment which is available upon request.

2.2. Acceptance testing

According to the regulations set by the EMREX User Group, every new NCP will have to undergo the acceptance testing in the development environment before being accepted for production. The development team should prepare their solution for such testing and contact the EMREX User Group to schedule the testing. URL of the new NCP will be entered into the development EMREG and the team will get access to the development environment installed by the EMREX User Group. There is also a fully functional test installation with SMP, NCP and EMREX client, with anonymized data from the database of the University of Warsaw. This testbed is described in the documentation mentioned in section 2.1.

Acceptance testing will cover various scenarios run by hand and also with the use of automatic tools.

It is recommended that the NCP source code will undergo an external review. If possible it should be posted in GitHub as open source.

A special attention should be given to analysis of transcripts of records coming from various countries. Such research gives information needed for the design of automatic conversion of the data sets provided by one institution into the data model of another institution.

2.3. Problems to solve, best practices, lessons learnt

When implementing the EMREX solution, the development team has to approach various issues and make decisions. Here are some comments which may help to design and implement the EMREX module for the local system.

1. The EMREX network is based on trust. Each country and HEI will use its own authentication method, but generally speaking the following seems to hold true:

   a. a federation of trust based on unique username and password can be used in the EMREX network,
b. A point of strong identification is needed, e.g. the student should be required to personally be present with a valid identity to get his credentials. This is however the routine procedure in higher education institutions.

2. The student using EMREX has to sign in twice – first in the local EMREX Client, then in the NCP. When the student is redirected back from NCP to the local system the system has to check if both user accounts are connected with the same physical person. The matching of names is proposed to be done with the Levenshtein method. Default threshold value is 10%, but each implementer needs to adjust it according to local needs. Due to limited added information and risk for allegations of discrimination, some partners have decided not to use the sex of the student as a parameter.

3. It is highly recommended that both SMP and NCP log information which can be used to track the behaviour of the Network. Some statistical information on the popularity of various NCPs may be needed to attract new partners. Personal information cannot be stored nor shared among partners. The EMREX User Group set some recommendations on the format of SMP and NCP log files which should be followed, if possible.

4. Legal issues constitute the important aspect of the EMREX Network. EMREX assumes that any student has the right to ask for his own achievement records. However, different legal rules can affect the strictness of the implementation, e.g.

   a. Is the achievement record public or private information? The answer to this question might affect the decision on how strong the authentication process should be;
   b. If the matching of names fails, what is the system allowed to do? Can the data be still transmitted to the partner with some warning or should the process be stopped?

5. Implementation of the EMREX Client and the NCP as part of the student portal is usually straightforward and takes a few weeks.

6. The challenging part of the implementation is the interpretation of student achievements – courses, grades and ECTS points – stored in the local database, and generation of the transcript of records as XML and PDF in the EMREX ELMO format.

7. Even more challenging is the reverse process – mapping the data on courses, grades, and credits obtained from the ELMO XML file into the tables of the local database.

8. The most challenging part is the change in institutional policies and European regulations. Neither a student nor an institution should be obliged to store or transfer paper versions of documents when electronic ones are available.

2.4. Automatic recognition

The ELMO XML format is the basis for the exchange of result information. ELMO is based on the CEN standard EN 15981-2011 EuroLMAI. EuroLMAI is a data model describing assessments, primarily Diplomas, Diploma Supplements and Transcripts of Records for higher educations. The schema describing the profile of the ELMO format used in EMREX is available in the EMREX GitHub repository.

The XML can contain the detailed information about the learner (student), the issuer (institution), the achievements (courses, grades, credits, working hours etc.) and in particular the PDF with the printable version of the transcript of records. It is highly recommended that the PDF embedded in the XML contains exactly the same information (that both elements are semantically equivalent). Be
Recommendations for developers

aware that in some cases that may not be true since some partners, during some transient period, may prefer to send PDF with the minimal required set of data in XML.

The important implementation decision is how the data carried in EMREX XML will be stored in the local database. It is possible to store it as one binary object and allow the users to view PDF. However, the more challenging solution is to unpack the data from XML and store them in the local tables, together with other student’s achievements. Such approach allows to use the data for automatic recognition, print them in the diploma supplement or use for any other purpose.

Some partners also send grade distribution for each course. Such distribution may be useful for grade conversion if no automatic method is available. The EMREX User Group plans to add to the EMREX scenario the grade conversion based on the methodology and tools supported by Egracons.

The design of the EMREX scenarios implemented in the local system should be driven by the local stakeholders. It is highly recommended to share experience with the partners from the EMREX User Group. Such scenarios should be simple and intuitive.

When importing data to the local system one has to take into account differences between sets of data delivered by partners. Partners may not send data which are required in the local system. The semantics of the transferred data must be clear to store the data in the local tables.

Problems to be solved include:

1. Differences between modules and courses.
2. Achievements sent in the context of the study program.
3. Course activities not always delivered.
4. Grade conversion may be necessary.

2.5. ELMO backward compatibility

The development teams should observe and take part in discussions on ELMO in GitHub since it is very probable that when the new institutions join the Network some change proposal will be issued.

Before issuing a change proposal, one has first to look at previous discussions to check if the proposal has not already been dealt with in the past.

The EMREX User Group will try to keep all changes backward-compatible but it may happen that some incompatibilities will be voted for by the ELMO society.

There are some unresolved issues in GitHub. The developers should review the list of unresolved issues and take part in the discussion.
3. Anticipation of potential risks

The evaluation of the project results was carried in a couple of diverse and complementary studies and not only delivered answers to questions concerning usability of EMREX for transfer of achievements and support of the mobility and recognition (see the Evaluation Report) but also revealed potential business risks for institutions incorporating EMREX into their daily processes. Such risks should be identified and appropriate management routines should be worked out. All new systems connected to the EMREX Network should be subject to the risk analysis. It should be performed by the owner of the system.

This chapter contains the list of potential risks that could be considered during the analysis. They are grouped into four categories. Their probability and impact are evaluated in the range from 1 to 5:

- **Probability:**
  - Lowest value is 1 – very low (never occurs).
  - Highest value is 5 – very high (occurs daily/always).

- **Impact**
  - Lowest value is 1 – can be ignored.
  - Highest value is 5 – very high/catastrophic.

### Availability risks

<table>
<thead>
<tr>
<th>Risk</th>
<th>Probability</th>
<th>Impact</th>
<th>Actions</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMREG downtime.</td>
<td>2</td>
<td>2</td>
<td>Redundancy and alerts.</td>
<td>EMREG downtime may not matter much as long as cached data can be used.</td>
</tr>
<tr>
<td>NCP downtime.</td>
<td>2</td>
<td>4</td>
<td>Redundancy and alerts.</td>
<td></td>
</tr>
<tr>
<td>EMREG is DDOSe for several days during diploma deadline.</td>
<td>2</td>
<td>2</td>
<td>Anti DDOS measures and caching.</td>
<td>EMREG data can be cached.</td>
</tr>
<tr>
<td>NCP is DDOSe. Unable to retrieve foreign diplomas from attacked country.</td>
<td>2</td>
<td>4</td>
<td>Anti DDOS measures.</td>
<td>Caching all diplomas is not an option</td>
</tr>
<tr>
<td>Unstable NCP which does not deliver continuous service.</td>
<td>2</td>
<td>4</td>
<td>Redundancy and alerts.</td>
<td></td>
</tr>
</tbody>
</table>
## Integrity risks

<table>
<thead>
<tr>
<th>Risk</th>
<th>Probability</th>
<th>Impact</th>
<th>Actions</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>ELMO data not compliant with student achievements.</td>
<td>3</td>
<td>5</td>
<td>Adopt good practices in the institution, test.</td>
<td></td>
</tr>
<tr>
<td>Different data sent in ELMO XML and in embedded PDF.</td>
<td>3</td>
<td>5</td>
<td>Adopt good practices in the institution, test.</td>
<td></td>
</tr>
<tr>
<td>Incorrect entries in the EMREG registry.</td>
<td>3</td>
<td>4</td>
<td>Acceptance testing of new partners, testing EMREG regularly.</td>
<td></td>
</tr>
</tbody>
</table>

## Correctness risks

<table>
<thead>
<tr>
<th>Risk</th>
<th>Probability</th>
<th>Impact</th>
<th>Actions</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>New HTTPS vulnerabilities are discovered. A hacker can replace NCP list data being downloaded from EMREG. This makes it possible to set up a fake NCP server and produce fraudulent diplomas.</td>
<td>4</td>
<td>5</td>
<td>Use correct server security configuration. Use cryptographic signatures on data, in addition to HTTPS.</td>
<td>Several critical SSL/TLS vulnerabilities have been found in recent years.</td>
</tr>
<tr>
<td>CA certificate leaked or abused.</td>
<td>3</td>
<td>5</td>
<td>Use signature on data in addition to HTTPS or use custom list of custom CAs.</td>
<td>Comodo, diginotar, Chinese CA</td>
</tr>
<tr>
<td>EMREG is hacked. SMP is communicating with correct server but receiving false data.</td>
<td>2</td>
<td>5</td>
<td>Pen-testing regularly and before major releases to secure application. Patch OS and third party software. Log changes made by users in application.</td>
<td></td>
</tr>
<tr>
<td>NCP is hacked. Places false data in EMREG.</td>
<td>2</td>
<td>5</td>
<td>Responsibility of each NCP but login solution (YubiKey etc.) can make abuse of EMREG more difficult.</td>
<td></td>
</tr>
<tr>
<td>Hostile code injected in SMP binaries (jar file).</td>
<td>2</td>
<td>5</td>
<td>Reproducible/deterministic builds.</td>
<td></td>
</tr>
<tr>
<td>Hostile code injected in mobility plugin source code Enabling forging diplomas and hacking systems using SMP.</td>
<td>2</td>
<td>5</td>
<td>Good practice, open source, code review.</td>
<td></td>
</tr>
</tbody>
</table>
### Recommendations for developers

| NCP sends data against user's wish. | 3  | 5  | NCP must ask for consent and make it very clear which data will be sent and where it will be sent. A white list of approved URLs for diploma delivery could be used to prevent fake web sites from receiving diplomas. White listing might not be an option if EMREX is to be used by recruiting agencies etc. |
| Open source. Makes it easy for hackers to find vulnerabilities. | 2  | 4  | Bug bounty program. |
| XML file does not validate against the ELMO schema. | 4  | 5  | Code review, testing. |

### Process/management risks

<table>
<thead>
<tr>
<th>Risk</th>
<th>Probability</th>
<th>Impact</th>
<th>Actions</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small number of participating countries.</td>
<td>5</td>
<td>4</td>
<td>Involve your mobility partners, be active in the EMREX User Group.</td>
<td></td>
</tr>
<tr>
<td>Changes in legal regulations.</td>
<td>3</td>
<td>3</td>
<td>Trace law on student achievements, data transfer between countries, privacy issues.</td>
<td></td>
</tr>
<tr>
<td>Process of registration of new NCPs lasting too long.</td>
<td>1</td>
<td>2</td>
<td>Be active in the EMREX User Group.</td>
<td></td>
</tr>
<tr>
<td>Expired student accounts.</td>
<td>5</td>
<td>5</td>
<td>Implement e-IDAS</td>
<td></td>
</tr>
<tr>
<td>The high cost of software maintenance.</td>
<td>4</td>
<td>3</td>
<td>Reuse software, hire companies recommended by partners with the experience in implementing EMREX.</td>
<td></td>
</tr>
<tr>
<td>Lack of consensus between partners if a student can choose grades or if failed courses should be transferred.</td>
<td>3</td>
<td>3</td>
<td>Be active in EMREX User Group, be active in GitHub.</td>
<td></td>
</tr>
<tr>
<td>Rogue employees.</td>
<td>1</td>
<td>5</td>
<td>Adopt good practices in the institution, organize training.</td>
<td></td>
</tr>
</tbody>
</table>