Title	**Models for the future organisation of technical R&D developments and training for metadata technicians (D8.2b)**
Work Package	WP8
Authors	Atle Alvheim, Ørnulf Risnes, Dag Kiberg, Håkon Nilsen
Source	WP8
Dissemination Level	PU (Public)

Summary/abstract

Future CESSDA Cooperation will be organised with one central administrative unit. There will be a need for technical development work on data **collection**, data **processing**, data **documentation** and data **storage**. Further, CESSDA is expected to run projects concerning semantic web technology, publishing to the web, facilitating data location and exploration over the Internet.

The required technological developments will be technically very sophisticated and cutting edge. The development work should be conducted in integrated teams. Three scenarios for the organisation of work are outlined. The cessda-ERIC as:

a. Developer;
b. Customer;
c. Co-ordinator.

Table of Contents

1. cessda-ERIC needs for technical development work

The proposed new legal entity (cessda-ERIC) will not be a European Data Archive itself. Operating in a distributed fashion across a number of clusters of expertise, ideally it should neither hold data nor disseminate data. The role should be that of a portal service, bringing together - in one central focal point - the tools and technologies needed to access what is maintained in a network of nodes across Europe. A key role will be in the development and implementation of standards, common tools and provision of strategic guidance. These latter tasks require access to some sort of funding, organized in a way that fits in and is in accordance with the overall organizational model of cessda-ERIC, as well as for the particular model for technical development work.

The cessda-ERIC is (so far) proposed to be set up with a hub and a number of members. The members consist of those bodies or States that fulfil the necessary requirements/qualify for membership (for States this will mainly be ministries/research councils). A member will be represented by a Resource Provider. Qualification for membership requires capacity, ability and willingness to comply with the organising rules and research promoting principles laid down and enforced by the central unit.

The new infrastructure is planned financed via a multi-level membership structure:

- Full Members (full financial commitment);
- Designate Members (only partial financial commitment);
- Affiliate Members (non or very little financial commitment).

The category of Designate Member is intended as a transitional rather than permanent status. Financial support from national ministries/research councils will be required at two levels:

- Directly funding to the cessda-ERIC hub;
- Funding of the operational Resource Provider organisations within the cessda-ERIC (fee).

This means that there will be collected fees from the members (may be not from the Affiliate Members). The members shall be liable for their respective contributions to the cessda-ERIC. In addition to collected fees it will probably follow some financial resources from the hub host country.

In order to have an effective and efficiency funding for procurement of technical development or solutions it is necessary that cessda-ERIC, represented by the hub, have some financial resources to use for technical development work or products. Basically, following the general funding scheme for cessda-ERIC there are two options for funding technical development:

- directly from the members and the hub host (i.e.. via the fee-system or ad hoc);
- from public funding (i.e. as project funding from EU).

It is unlikely to see a scenario including raising loans or for instance inviting external investors.

2. cessda-ERIC Technical Tasks and Development Projects

The cessda-ERIC will be involved in all aspects of running an integrated data service for social science research. This will mainly be focused on setting standards and giving support for:

a) New data collection, supplying know how and tools for data collection for different types of data and in different contexts, given the need for support;
b) Collecting, organising and archiving data already available from a variety of sources, but deemed valuable and worthy of long term archiving;
c) Data documentation work, leading up to publication and archiving processes, in particular development of models, tools and best practices;
d) Web-publication of data, with retrieval and security systems, to promote the comparative perspectives and help solve communication problems across a linguistically dispersed Europe;
e) Facilitation of user interfaces, user location and processing processes, data access.

This requires a succession of technical development projects. The work in this field is fast moving, although in its infancy. Much CESSDA development work will have the character of being leading edge semantic web development.

Particularly demanding will be problems concerning identification and integrity of resources and resource systems in an Internet setting and linking of the traditional archive and library domains.

Basically, this kind of software and systems development should be conducted in balanced teams, spanning many types of specialist qualifications, by no means only technical. As these are highly sophisticated problems, leading edge qualifications are needed in a number of areas:

a) Accurate analysis and descriptions of user community needs;
b) Specification writing, converting user problems into technical tasks;
c) Network specialists, architecture;
d) Database specialists, data storage, long term archiving organisation, data preservation;
e) Programmers, application developers.

Technical developments include, but are not necessarily limited to, the following aspects:

a) Architecture developments;
b) Software development;
c) Hosting/maintenance;
d) Technical support;
e) Testing/documentation.

3. Organisational scenarios for technical development

We recommend that all aspects regarding the technical infrastructure should be led and coordinated by the central ERIC, supported by an expert group (or "technical committee") with representatives from the different archives. It is believed that such an expert group will be crucial to the success of the coordination efforts. As the envisioned infrastructure aims at creating an interoperable network from largely heterogeneous and autonomous services deployed at the member archive-sites, the task of coordinating the interoperability infrastructure will be a comprehensive one. The sheer number of cooperating services represents in itself a challenge in terms of estimated workload required for coordination; the fact that the heterogeneity between member-archives is very high multiplies complexity in interoperability efforts. Participation from archives in the form of an expert group is therefore highly recommended.

However, technology is only part of the question. There will always be an additional question on the purpose of the development. In the envisioned CESSDA infrastructure there are two major user interests. One is the data archives that work on data storage, preservation and delivery. The other is the researchers that participate the most heavily at each end of the process, in data collection and data analysis. There is a need to represent both these user interests in the envisioned expert groups.

For technical development in such an organisational setting, we have chosen to contrast three different organisational models:

3.1. ERIC as developer:

In this model, the ERIC-hub is tightly coupled to a development department/team, which is led by a responsible person within ERIC. This model is similar to what is the present reality in archives with development departments. Preferably, the team should be co-located with ERIC – but this model can also work with a more distributed team, where team members sit in different archives. A loosely coupled, distributed team-setup would lack some benefits compared to a co-located one, whereas it could be easier to recruit members to such a team from existing archives if they can remain at their home institution.

Regardless of co-located or distributed structure, an ERIC-development team should meet the following recommendations:

Team composition:

- 5-10 developers, 1-2 program managers (who report to technical ERIC-manager and coordinate team efforts on a day-to-day basis).

Qualifications:

A range of qualifications are needed, in e.g. the following areas:

- Service Oriented Architecture;
- Web-design and UI-design;
- Database management systems/Database administration;
- Security architecture and implementation;
- XML-processing, metadata handling;
- Scalable solutions (load-balancing, clustering, fail-over, mirrors, backups).

Pros:

- Tight co-located team;
- Close relationship between development team and ERIC.

Cons:

- High running costs;
- Single point of failure (i.e. if centralized development fails the entire service network fails) compared to distributed development.

3.2. ERIC as customer:

In this scenario, the ERIC acts as a customer, issuing tenders for specific development(s), hosting and supporting contracts.

The tender-process could either be closed (i.e. open to CESSDA member archives only), open (i.e. open to all interested parties) or semi-open (i.e. open to selected parties from within or outside the CESSDA-environment).

Pros:

- By tendering the work, the cessda-ERIC puts itself in the position of the "paying customer". Normally, a clear customer/supplier relationship will create desired dynamics to the development process. Note that this is highly dependent on ERIC's "purchasing competence" and its ability to act as a professional "product owner" in the different projects;
- By tendering the work, the cessda-ERIC hands over to the suppliers to work and utilize the appropriate methodologies for achieving the well-defined result;
- Tendering usually creates competition between interested parties, and more alternative approaches are likely to emerge during a tendering process.

Cons:

- Potential conflict of interest; as we suggest that the ERIC is supported by an expert group with participants from archives that are also potential suppliers, legal or political problems could emerge in a scenario like this;
- Requirements of the final solution must be very well understood, and part of the issued tender. While understanding requirements up front is demanding but valuable, modern software development methods (e.g. SCRUM, Lean) suggest that requirements and solutions are defined between customer and supplier in a dynamic manner throughout the project period. Investigation into how a tendering approach can be combined with agile development methods is needed;
- Potential bureaucratic overhead in the tendering process.

3.3. ERIC as mere coordinator:

This is the model closest to the current development model in CESSDA.

In this model, the cessda-ERIC would coordinate work between independent archive teams, and possibly (but not necessarily) allocate funding to the said teams. In this model, and especially if development is funded by the archives themselves, the development teams would not report directly to the cessda-ERIC, but to their local archive.

Due to the lack of control that the ERIC will have over the development process, it is difficult to recommend this kind of model. The ambitions in current technological developments will require far tighter management and reporting than possible in such a model.

However in saying that, a model like this could potentially work complementary to one of the others. The reasoning behind this statement is that many archives will keep in-house development teams that create software for archiving purposes. If the development of this type of non-mission-critical software could be coordinated between CESSDA-members, duplicated efforts could be avoided, and the pool of software available to CESSDA-members could expand in the most efficient manner. Note that IPR-agreements need to be worked out for such a model to perform optimally.

Pros:

- Could be a good complementary model to one of the others.

Cons:

- Unclear authority dynamics (i.e. ERIC is without customer-leverage);
- Conflict of interest; if tasks are assigned to member archives without competition (comparable to that of the tender-approach) could lead to political issues;
- Unclear ownership/IPR of the outcome?

4. Decentralized API-driven team development

Abstract

Development of software solutions in co-located teams is the most common approach today, since it provides the easiest way of communicating with each member of the team. Software development over a distance is hard even for experienced teams. Still, more and more projects are scattered over a vast geographical area, and can result in a problematic and cumbersome development process, often resulting in less than enthusiastic teams and a solution that neither parties will be satisfied with.

Industrial engineering has a number of established methods for development and production of goods, from small companies with only a few employees to large construction sites with tens of thousands of workers all performing different tasks and meeting the demands specified within the set time schedule.

In this regard, software development is still in its infancy. Continuous production for co-located teams is nearly impossible without a good foundation to base the solution on, and development across a distance will not help in this matter.

API-driven team development is a way to divide the project into stages, where each team has a specified, predefined task in a series of tasks, which all will be based on one or several sets of API's. A single API can provide the basis for several teams to start developing simultaneously, or each team can chronologically inherit API's of other teams.

Though hard to accomplish, API-driven development is possible even at a distance. As long as each team has a defined role and is able to deliver or contribute to the API, other teams will have the tools to complete their part of the project.

Decentralized API-driven team development requires a fixed set of tasks before any teams start the actual programming. Through a preparatory phase or prototyping through a committee of team members, the project description should involve every task that is required for each team to complete their part.