D4.1

Report on risk management for e-Infrastructures

Document information

Title	Report on risk management for e-Infrastructures
ID	CLARINPLUS-D4.1 (CE-2016-0743)
Author(s)	Franciska de Jong, Erhard Hinrichs, Dieter Van Uytvanck, Claus Zinn
Responsible WP leader	Erhard Hinrichs
Contractual Delivery Date	2016-03-01
Actual Delivery Date	2016-03-31
Distribution	Public
Document status in workplan	Deliverable

Project information

Project name	CLARIN-PLUS
Project number	676529
Call	H2020-INFRADEV-1-2015-1
Duration	2015-09-01 – 2017-08-31
Website	www.clarin.eu
Contact address	contact-clarinplus@clarin.eu

Table of contents

1 Executive Summary 2

2 Introduction 3

3 Identified risks 4

3.1 Timing 4

3.1.1 Possible Causes 4

3.1.2 Preventive Actions 4

3.1.3 Contingency Plans 4

3.2 New Practices and Paradigms 5

3.2.1 Preventive Actions 5

3.2.2 Contingency Plans 5

3.3 Critical Mass of Scholars 5

3.3.1 Preventive Actions 5

3.3.2 Contingency Plan 5

3.4 Funding 6

3.4.1 Possible Causes 6

3.4.2 Preventive Actions 6

3.4.3 Contingency Plans 6

3.5 Cooperation 6

3.5.1 Possible causes 6

3.5.2 Preventive actions 6

3.5.3 Contingency plans 7

3.6 E-Infrastructure Risks Outside the Competence of Humanities 7

3.6.1 Availability Issues 7

3.6.2 Data Loss 8

3.6.3 Security Incidents 8

3.6.4 Abuse, IPR infringement, falsification and complaints 9

3.6.5 Human Resources 9

4 Dissemination and further use of this risk catalogue 10

5 Conclusion 10

References 11

1 Executive Summary

In August 2013 the European Commission has published a report entitled "*Assessing the projects on the ESFRI roadmap, A high level expert group report*". The report summarizes the assessment results for thirty-five research infrastructure initiatives that were included on the ESFRI roadmap at that time. The scope of the assessment included six modules: cost and financial structure; governance and legal structure; stakeholder engagement and financial commitments; human resources and project management; user strategy and risk.

For the CLARIN research infrastructure, the report issued a recommendation to further develop a strategy for risk analysis and management. The analysis focuses on risks related to e-Infrastructure. The purpose of the present deliverable is to address the report's recommendation in detail.

The deliverable identifies six risks with respect to Timing, New Practises and Paradigm, Critical Mass of Scholars, Funding, Cooperation, and E-infrastructure Risks Outside the Competence of Humanities. For each risk, this analysis identifies possible causes, suggests preventive actions, and also outlines contingency plans in case the preventive actions were to fail.

2 Introduction

On October 1, 2013 the EC published a report entitled "*Assessing the projects on the ESFRI roadmap, A high level expert group report*". (henceforth referred to as the AEG report[1]). The AEG report summarizes the assessment results of the Assessment Expert Group (AEG) for thirty-five research infrastructure initiatives that were included on the ESFRI roadmap at that time. The scope of the AEG assessment included six modules: cost and financial structure; governance and legal structure; stakeholder engagement and financial commitments; human resources and project management; and user strategy and risk.

For the CLARIN research infrastructure, the AEG issued a total of thirteen recommendations, one of which concerns risk analysis and management, and which reads as follows: *Risks analysis and management should be developed further. Risks outside the competence of the Humanities - risks related to the e-infrastructure (crashing of servers, virus infection, hacking and stealing of data, falsification and integrity of data, introduction of ethically and personally offending papers, contract with providers) should be elaborated more explicitly.* (AEG report, p. 14). The purpose of the present deliverable is to address this recommendation in detail.

At the request of its General Assembly, CLARIN ERIC had addressed the issue of risk analysis and management already prior to and in parallel with the recommendations of the AEG report. The Board of Directors of CLARIN ERIC took the lead in this matter and sought continued input from all national consortia as well as from the national stakeholders. This resulted in a risk analysis document that has become an integral part of CLARIN's Strategic Plan (document CE-2013-0238). The document was discussed by the CLARIN ERIC Advisory Board and subsequently approved by the General Assembly of CLARIN ERIC. This assessment of risks and how to deal with them, which was vetted by all relevant CLARIN committees and by the General Assembly of CLARIN ERIC, forms the basis of the present deliverable. In keeping with the specific recommendation of the AEG cited above, this deliverable puts special emphasis on the risks related to the e-infrastructure, which are addressed in detail in section 3.6.

The remainder of this deliverable is structured as follows: Section 3 addresses a total of six risks that pertain to the following aspects: timing (section 3.1); new practises and paradigms (3.2); critical mass of scholars (3.3); funding (3.4); and cooperation (3.5). The subsections 3.1 – 3.5 summarize the risk assessment that was conducted by CLARIN ERIC independently of the AEG report. Section 3.6. goes beyond the scope of this earlier assessment and explicitly addresses e-Infrastructure risks outside the competence of the Humanities. Each of the six subsections describes the risk in question, identifies possible causes, suggests preventive actions, and formulates contingency plans in case the preventive actions were to fail.

[1] https://ec.europa.eu/research/infrastructures/pdf/jd-final-aegreport-23sept13.pdf

3 Identified risks

This section includes an extended version of the risk assessment made earlier by CLARIN ERIC as an addendum to its Strategic Plan (CE-2015-0657), which was also submitted to the AEG.
With the resources available through CLARIN-PLUS we added a similar analysis with regards to typical e-infrastructure risks, which is provided in the last section.

3.1 Timing

The implementation of common CLARIN ERIC infrastructure is delayed or stalled, due to technical or logistical reasons.

3.1.1 Possible Causes

1. Lack of technical expertise at the national level.
2. The pre-conditions necessary for technical integration at the European level are missing at the national level.
3. Lack of standardized data formats prevents interoperability of resources and tools.
4. Legal obstacles prevent (trans-national) access to resources and tools.
5. Lack of Communication between the National Consortia

3.1.2 Preventive Actions

At any point in time there will be an active integration plan with explicit targets, aiming at integrating all members, integrating between the technical and the content level, integrating standards, legal aspects *etc*. The integration activities will comprise workshops, documents, conferences *etc*. Committees and working groups have been created, in particular the Standing Committee for CLARIN Technical Centres, the Centre Assessment Committee, the Standards Committee, and the Legal Committee. (This is already part of the strategy and involves the pillars *Legal, Integration of data, Integration of services,* and *Preservation*.)

In addition, the CLARIN ERIC secretariat will make sure to provide all existing specification documents (many created in the preparatory phase) in a well-structured and easily accessible way.

In order to guard against cause 5, CLARIN ERIC has put in place the National Co-ordinators Forum, which convenes the scientific coordinators from the ERIC members and ensures direct and continuous communication among all national consortia.

3.1.3 Contingency Plans

If technical pre-requisites cannot be solved at the national level, load-sharing measures by other CLARIN members need to be considered and, if possible, implemented as temporary workarounds.

If integration steps and milestones specified in the actual integration plans are not reached in a timely fashion, CLARIN ERIC's Board of Directors will consult with the National Coordinators to identify the causes for these delays, will work with the National Coordinators on overcoming these obstacles, and will revise the work programme in consultation with the National Coordinators, if necessary.

If communication within and across Working Groups, Committees, General Assembly, and National Coordinators Forum is insufficient, CLARIN ERIC's Board of Directors will ensure adequate information flow, by contacting the relevant Chairs and Vice-chairs and by calling face-to-face or virtual meetings with the relevant groups if necessary on short-notice.

3.2 New Practices and Paradigms

The implementation of the common CLARIN ERIC infrastructure does not conform to the emerging state-of-the-art technology for research infrastructures (RIs), or it does not take into account the best practises followed by other RIs.

Consider, for instance, CLARIN's Authentication and Authorisation system, which is based on Shibboleth, and which has been or is being implemented in all CLARIN countries. Shibboleth may turn out to be too heavy, and other more lightweight systems may become available. In this case, CLARIN will need to consider adopting the new system.

3.2.1 Preventive Actions

CLARIN has taken up a leading role in other European projects and initiatives such as EUDAT and DASISH to actively shape, help define, and further develop the state-of-the-art for RIs in the Humanities and Social Sciences. CLARIN is also actively participating in the Research Data Alliance (pillar *Preservation*).

CLARIN ERIC's Board of Directors and CLARIN ERIC Working Groups and Committees regularly participate in high-profile events organized by other RIs and participate in Working Groups of Standards Organizations such as ISO or TEI.

3.2.2 Contingency Plans

External assessment is needed to detect important areas where CLARIN does not follow the development of state-of-the-art technology and best practises, and where it is not able to respond adequately to user needs. The Scientific Advisory Board has been established in order to avoid this situation.

3.3 Critical Mass of Scholars

CLARIN ERIC's infrastructure does not attract a critical mass of scholars in the Humanities and Social Sciences.

3.3.1 Preventive Actions

A critical mass of scholars will mainly be attracted through plans for the pillars *Ease of access*, *Integration of data* and *Integration of services*.

It is paramount to ensure an easy access to all CLARIN technology and services. Technical obstacles will be minimised, and users will need to be supported by a help desk and adequate documentation.

To attract new scholars, CLARIN will showcase its infrastructure during conferences, workshops, courses, and other knowledge sharing events.

At the European level, CLARIN will identify and respond to appropriate funding opportunities for the easing of trans-national access to the CLARIN infrastructure. At the national level, CLARIN will seek close and continued interaction with all scientific communities of the Humanities and Social Sciences.

3.3.2 Contingency Plan

CLARIN ERIC will team up with other infrastructures close to the humanities to use joint communication and dissemination channels.

3.4 Funding

If CLARIN would get into a situation where no national funding remains, this would threaten both the finalization of the implementation as foreseen, as well as the sustainability of the data repositories and services that are already up and running.

3.4.1 Possible Causes

As there is a relatively lean budget at the European level, funding for the Implementation Phase of European RIs relies largely on national funds. This also holds for CLARIN ERIC, which funds itself through membership fees. In CLARIN, there is no common timetable that synchronises activities across national CLARIN consortia. This includes the decision making process with respect to national funding.

3.4.2 Preventive Actions

This risk is addressed through growth of the CLARIN membership base and a detailed sustainability strategy. Details for the latter will be provided in CLARINPLUS-D6.2 (Financial and organisational sustainability).

3.4.3 Contingency Plans

If national funding for a CLARIN member would dry out, it will be attempted to transfer the services and content of repositories of the member to institutions (within or outside the member country in question) with secured funding.

If no CLARIN member state can be identified to take over the data and/or services in question, which we consider highly unlikely, it will be attempted to transfer data to another research infrastructure to preserve the existing resources in question. However, the helping research infrastructure will most likely be unable to also take over or further develop the services in question (lack of expertise).

CLARIN ERIC will never run out of funding at once. The members always allocate their contributions for a number of years, typically 3-5 years, sometimes with a gap in between two periods. As a result, a drastic budget reduction would become evident within a 2 to 3 years notice period. This will give CLARIN members a sufficient amount of time to negotiate alternatives for securing the research data and services. Of course, this would require an adjustment of the governance of the infrastructure. In a period without funding the role and responsibilities of CLARIN ERIC would be very different from what they are now.

3.5 Cooperation

The implementation of the common CLARIN infrastructure does not reach its full potential or is even damaged by a lack of cooperation among national consortia.

3.5.1 Possible causes

The national consortia rely on national funding and need to serve their funding agencies.

3.5.2 Preventive actions

It is an integral part of the CLARIN ERIC governance structure to prevent this. In particular, the National Coordinators Forum and the Standing Committee for CLARIN Technical Centres have been established for this purpose.

3.5.3 Contingency plans

If the organisational structure of CLARIN ERIC is not sufficient, then the ERIC will have to re-assess the structure, including the assessment of the light governance structure. Probably the best solution would be to have more power and more resources in the central part of the budget, which would require to raise the annual fee of member states.

3.6 E-Infrastructure Risks Outside the Competence of Humanities

Next to the risks mentioned before, CLARIN should also take into account the more general risks related to using and providing e-infrastructure services, as explicitly mentioned in the AEG report.

In the wider context of Distributed IT-Infrastructure we refer to the paper *A Trust Framework for Security Collaboration among Infrastructures* from Kelsey et al. (2013), which provides a good overview of such risks. The paper formed the EUDAT context for risk analysis (work package 6, operations). The list of e-Infrastructure issues below has also been inspired by Kelsey et al. (2013).

It should be noted that each of the risks apply both to the distributed managed services (provided by individual CLARIN centres) and to the central services (provided by CLARIN ERIC). It will be important to strike the right balance between addressing the issues at a global and local level.

Although the services provided via the centres are not under direct control of CLARIN ERIC, the centre assessment (and the Data Seal of Approval self-assessment) should guarantee that the required prevention and contingency measures are in place.

3.6.1 Availability Issues

An online service provided by CLARIN is not reachable or usable.

Possible causes	**Preventive actions**	**Contingency plans**
Crashing of servers, due to failing hardware.	Redundant hardware with failover functionality Virtualisation with failover functionality	Fall-back plan to migrate services to other physical server
Crashing of servers, due to failing third-party software (e.g. Operating System, middleware, ...)	Software test plans Using proven software Redundant software setup	Roll-back to previous stable version after update
Crashing of servers, due to failing self-developed software.	Software test plans Unit tests, Continuous Integration	Roll-back to previous stable version after update Fast-response debugging
(Distributed) Denial of Service attacks	Redundant hardware & software setup Monitoring resource use, in collaboration with hosting providers	Blacklisting of IP-ranges

In each case, the contingency plan will be reinforced by a transparent and fast communication about the issue at stake, via:

- The www.clarin.eu/status service monitor page
- Mails to all CLARIN centres via the all-centres mailing lists

3.6.2 Data Loss

Data stored within CLARIN is lost.

Possible causes	**Preventive actions**	**Contingency plans**
Hardware failure	Redundant hardware with failover functionality (e.g. RAID-based storage) Regular hardware replacement	Restore (local) backup
Hardware failure due to major incidents, e.g. fire or flooding	Server-room protection measures	Restore (off-site) backup
Software bugs	Software testing Peer review	Restore backup
Human error	Raising awareness, using software design patterns that make it harder to make major mistakes	Restore backup

3.6.3 Security Incidents

A non-authorised person or system obtains access to a CLARIN server or application.

Possible causes	**Preventive actions**	**Contingency plans**
Benevolent attacks (ethical hacking)	None. Inviting ethical hackers to test the server security could be helpful to identify issues.	Learn from identified issues and apply improvements
Malicious attacks (cracking)	Keeping software up-to-date Standard security best practices (firewall, encryption, password policies) Run regular checks (port scans etc.)	Follow the security incident handling procedure
Viruses, malware and trojans	Use only trustworthy software sources (*e.g.*	Follow the security incident handling procedure

	official Linux distro repositories) Run regular scans	
Software bugs and security vulnerabilities	Software testing Peer review Use standardized and well-maintained security libraries and packages	Follow the security incident handling procedure
Human errors	Raising awareness, using software design patterns that make it harder to make major mistakes (e.g. restrictive default access policies)	Follow the security incident handling procedure

An exact procedure to handle security incidents will be developed according to the guidelines put down in the Handbook for Computer Security Incident Response Teams (CSIRT) by West-Brown et al. (2003) and in close collaboration with EUDAT.

3.6.4 Abuse, IPR infringement, falsification and complaints

Inappropriate data is published via CLARIN, e.g. by uploading it into a repository.

Possible causes	**Preventive actions**	**Contingency plans**
Malicious data is published (spam, viruses, …)	Regular manual content checking procedures Require a high trust level and traceability for uploading data	Remove the malicious data
Data that infringes intellectual property rights is published	Request the specification of a detailed license when uploading data Avoid liability via appropriate terms of use	Careful consideration of the case and interaction with a to be instantiated ethical committee in case of doubt. In case of proven infringement, remove the data.
Forged data is published	Avoid liability via appropriate terms of use	Careful consideration of the case and interaction with a to be instantiated ethical committee in case of doubt. In case of proven forgery, remove the data.
Complaints about published (meta)data for a reason not listed above (e.g. reputation damage)		Careful consideration of the case and interaction with a to be instantiated ethical committee in case of doubt.

3.6.5 Human Resources

Possible causes	**Preventive actions**	**Contingency plans**
Temporary lack of know-how due to personnel absence	Ensure know-how is documented and minimally shared among 2 persons Keep extensive and up-to-date documentation Good planning of vacation periods	Postpone activities with low or medium priority
Permanent loss of know-how due to personnel changing jobs	Ensure know-how is minimally shared among 2 persons Keep extensive and up-to-date documentation Provide attractive working conditions	Timely knowledge transfer to other personnel members

4 Dissemination and further use of this risk catalogue

CLARIN ERIC will build on the information gathered for this document and to update it over time. The e-Infrastructure risk assessment section will be distributed to the members of the Standing Committee for CLARIN Technical Centres to gather additional feedback and will then be further maintained at the CLARIN developer wiki, to provide an up-to-date overview for those new to the CLARIN infrastructure.

5 Conclusion

The deliverable identifies six risks with respect to Timing, New Practises and Paradigm, Critical Mass of Scholars, Funding, Cooperation, and E-infrastructure Risks Outside the Competence of Humanities. For each risk, the risk analysis identifies possible causes, suggests preventive actions, and also outlines contingency plans.

The document gives a concise account of all identified risks and how to contain them. The document will be distributed to all relevant CLARIN stakeholders, and actively consulted and maintained at regular intervals during the duration of the CLARIN-PLUS project.

References

Calvia-Goetz, A. (Chair), Franciosi, A., Larsen, S., Marks, J., Tichmann, K., Wade, R., Zic Fuchs, M. (2013). *Assessing the projects on the ESFRI roadmap – A high level expert group report.* European Commission, Directorate-General for Research and Innovation, Directorate B – European Research Area Unit B3 – Research Infrastructures.

Kelsey, D., Marsteller, J., Kaila, U., Kanellopoulos, C., Chadwick, K., & Gaines, I. (2013). *A Trust Framework for Security Collaboration among Infrastructures* (No. FERMILAB-CONF-13-143-CD, p. 011). Retrieved (March 15, 2016) from the website: http://pos.sissa.it/archive/conferences/179/011/ISGC%202013_011.pdf

West-Brown, M. J., Stikvoort, D., Kossakowski, K. P., Killcrece, G., & Ruefle, R. (2003). *Handbook for computer security incident response teams (csirts)* (No. CMU/SEI-2003-HB-002). Retrieved (March 15, 2016) from: http://www.sei.cmu.edu/reports/03hb002.pdf