Document information Summary

Date	27.03.2017
Document title:	D14.4 Report on Use Cases, Requirements, Metadata and Interoperability of WP 14
Leader Partner	IG PAS
Main Author(s):	Mariusz Sterzel
Contributing author(s):	Karolina Chodzińska and EPOS-IP WP14 Team
Reviewer(s):	Beata Orlecka-Sikora
Approved by:	
Target audiences:	Project partnership, European Commission
Keywords:	Implementation phase, EPOS-IP, Interoperability,
Deliverable nature:	External
Dissemination level:	Public
Delivery date:	M18
Version:	01

TABLE OF CONTENTS

SUMMARY

This document summarises status of the service provided by WP14: Thematic Core Service Anthropogenic Hazards platform. The report includes information concerning the level of development of the TCS AH, including:

- Status of TCS-ICS integration and interoperability with other TCSs,
- Status of AH applications integration,
- Status of AH Episodes integration.
- Integration plans until the end of EPOS IP Project.

The overall level of integration advancement is in line with WP14 development plan. There were no major issues concerning TCS AH service development. All the harmonisation and TCS AH – ICS-C interoperability matters are worked out in close collaboration with EPOS WP6 and WP7.

1. Introduction

EPOS THEMATIC CORE SERVICE ANTHROPOGENIC HAZARDS, further on referred to as TCS AH aims to integrate distributed research infrastructures (RI) to facilitate and stimulate research in the area of anthropogenic hazards (AH), especially those associated with the exploration and exploitation of geo- resources. To fully describe anthropogenic phenomena one cannot focus on one type of data only. In contrast to other TCSs AH requires sets of data coming out of several sources. Thus, by necessity those are provided for the user as a whole and named episode. So, by design, the episode is a comprehensive data description of a geophysical (e.g. deformation) process, induced or triggered by technological activity, which under certain circumstances can become hazardous for people, infrastructure and the environment. The episode consists of a time-correlated collection of geophysical data representing the geophysical process, technological data representing the technological activity, which is the cause of this process and all other relevant geo-data

describing the environment, in which the technological activity and its result or by-product, the geophysical process take place. Episodes are enriched by applications, uniquely designed for the discrimination and analysis of correlations between technology, geophysical response and resulting hazard. Bound together within an IT environment enables for easy access to the episodes' data, analysis via integrated software and advanced visualisation od data and their products.

The prototype of the platform has been already build within the frame of project "IS-EPOS: Digital Research Space for EPOS Purposes". The project was accepted for co-financing by the European Regional Development Fund as part of the Operational Programme Innovative Economy. The IS-EPOS project aimed to build a prototype of TCS Anthropogenic Hazards belonging to pan-European multidisciplinary research platform created within EPOS infrastructure. The prototype deals with anthropogenic seismicity. The IS-EPOS IT platform integrates currently six episodes of anthropogenic seismicity linked to:

- Underground hard rock and coal mining in Poland
- Hydro energy production in Poland and Vietnam
- Geothermal energy production experiment in Germany.

The platform built upon the InSilicoLab technology by Cyfronet, provides low level software services for data browse, selection and visualisation. Those are accompanied by a number of high level applications and tools for advanced data processing and analysis. It has been assumed that the prototype will have all the basic functionality of the target TCS Anthropogenic Hazards service and that it will be developed to its target form during the implementation phase of EPOS. In parallel to work undertaken within IS-EPOS project, Cyfronet team joined EPOS PP working group 7 - e-infrastructures and virtual communities. This cooperation continues to achieve full interoperability between TCS-AH prototype and the EPOS infrastructure.
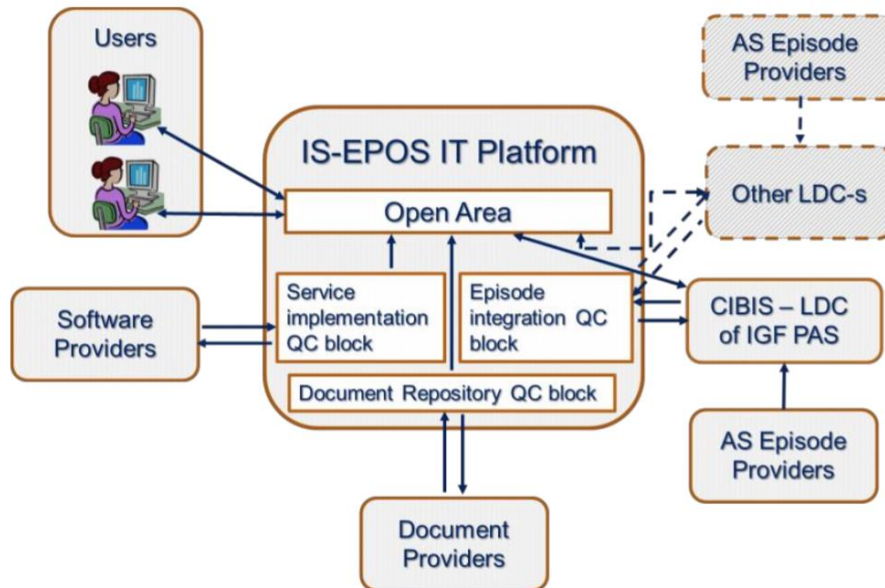
Picture 1:Topology of IS-EPOS platform – a prototype of TCS AH.

As mentioned six episodes are already integrated into IS-EPOS Platform. Within EPOS IP project TCS AH will integrate next 20+ AH episodes, connected with conventional and unconventional hydrocarbon extraction, reservoir treatment, underground mining and geothermal energy production.

From the architecture point of view TCS AH prototype is built of two major components: the AH local data centres (CIBIS, LDC), gathering episodes and their multidisciplinary data and the IT platform binding together access to the episodes' data, software and services. Platform provides also access to computational resources for advanced analysis and visualisation. All from users' browser.

The implemented application and tools will comprise the services to ensure full interoperability with EPOS Integrated Core Service, ICS-C, the services to secure communication and data transfer from local data centres to the central hub, the specialized bespoke applications for studying AH issues and the new generic and episode dedicated visualization engines.

The construction plan assumes that TCS AH IT environment will become a virtual laboratory equipped with tools for intercommunity collaborative research, will facilitate training in the field and will be also a source of competent information about anthropogenic geophysical hazards.

The access to most of the episodes integrated into TCS AH is and will remain open for all. For statistical purposes visitors of TCS AH register but no authorization is required. However, an authorized user, can reach higher level processing functionalities of TCS AH. More information about users' access and data policies can be found in deliverable D14.1 Report on Legal and Governance Framework for TCS Implementation of WP14
.

2. Priority List of DDSS

The only service provided by Anthropogenic Hazards Thematic Core Service is the TCS AH platform built upon IS-EPOS science gateway/virtual laboratory. However due to platform complexity and vast majority of different applications, episodes and tools available, it is convenient to divide them into separate classes.

2.1.   Applications

Class "applications" includes a portfolio of AH software integrated with the platform. To assure flawless utilisation by users each scientific code is unified to operate on data formats available in the platform. The outputs generated by the applications have been modified accordingly, to integrate with TCS AH platform visualisation engines. The following list of the applications has been scheduled for the integration with the TCS AH platform. If not otherwise states the integration of a given application is scheduled for M1-M24.

Table 1: Task 14.3 applications

To be implemented				Implementation mode/ plan of technical implementation
Data	Data Products	Software	Services	
			Modified TCS-AH authorization allowing for new scheme of roles	Under implementation
			Service for Integration TCS AH with ICS authentication and authorization	Implemented
			Developed integration with new episodes and open data	Under implementation. See section Episodes for details.
			New workspace functionalities	Under implementation
			New type of visualizations (specific types will be defined while integrating other planned DDSSes)	Under implementation. This heavily depend on integrated applications and episodes.

Table 2: Task 14.4 applications

To be implemented				Implementation mode/ plan of technical implementation
Data	Data Products	Software	Services	
AH episode data heterogeneity and availability assessment at external partners				The episode data validation is performed before the episode set is opened to the TCS AH users.
AH episode data initial standardization at external partners following established guidelines.		Ad-hoc conversion tools to homogenize AH episode data in future Local Data Centers (LDCs)		Partially implemented for data formats accepted for the platform. The technological/ production data usually require special tools which are created when necessarily
AH episode data upload validation and QC (standards, formats, completeness) in LDCs		Task management system for quality control (QC) and management towards efficient data integration into TCS AH		Operational
Preparation of metadata according to TCS metadata scheme		Management system of metadata in Local Data Centers (CIBIS)		Operational
TCS AH – ICS-C metadata integration		Metadata integration with ICS-C		Integration pending
Contextual analysis and QC of AH episode data, metadata generation				Operational

Table 3: Task 14.5 applications

Data	Data Products	Software	Applications	Implementation mode/ plan of technical implementation
			GSS1: Effective stress drop estimate	Integration pending
			GSS2: Moment tensor	Integration pending
			GSS3: Location of emergent – sustained signals	Integration pending
			GSS4: Tracking changes in elastic parameters	Integration pending
			GMS1: Fracture Network Models (mechanical stress)	Integrated
			GMS2: Fracture Network Models (poroelastic stress)	scheduled for second half of the 2017
			GMS2: Poroelastic Rate Proxies	scheduled for second half of the 2017
			GMS3: Stress and strain changes	scheduled for second half of the 2017
			QIS1: seismicity patterns tested against point process models	scheduled for second half of the 2017
			QIS2 seismicity patterns tested against reshuffled catalogues	scheduled for second half of the 2017
			A Template-Matching based Detection Algorithm	Integration pending

Table 4: Task 14.6 applications

Data	Data Products	Software	Applications	Implementation mode/ plan of technical implementation
			14.6.1: Testing stationarity	Integrated
			14.6.2: Correlation analysis: • Autocorrelation • Cross correlation • Coefficient of randomness • Correlation Analysis: Priestley-Subba Rao Test	Integrated
			14.6.3: Estimation of source parameters in time-varying production parameters geometry	Integration pending
			14.6.4: Estimation of time-changeable hazard parameters	Integrated
			14.6.5: Modelling uncertainties and interval estimation of hazard parameters	Integration scheduled for M25-40
			14.6.6: Estimate of maximum possible magnitude from induced deformation	Integration scheduled for M25-40
			14.6.7: Ground Motion Prediction Equations	Integration scheduled for M25-40

Table 5: Task 14.7 applications

To be implemented				Implementation mode/ plan of technical implementation
Data	Data Products	Software	Applications	
		*	Merger core application	Integration pending
*	Database with generic data			Integration pending
*	Database structure (for specific project data)			Integration pending
			Specific IAMs	Integration scheduled for M25-40

2.2. Episodes

As already mentioned in mentioned in the Introduction to fully describe anthropogenic phenomena one cannot focus on one type of data only. In contrast to other TCSs AH requires sets of data coming out of several sources. Thus, by necessity those are provided for the user as episodes. So, by design, the episode is a comprehensive data description of a geophysical (e.g. deformation) process, induced or triggered by technological activity, which under certain circumstances can become hazardous for people, infrastructure and the environment. The episode consists of a time-correlated collection of geophysical data representing the geophysical process, technological data representing the technological activity, which is the cause of this process and all other relevant geo-data describing the environment, in which the technological activity and its result or by-product, the geophysical process take place. Detailed list of the episodes and their integration status can be found in Table 6: Episode list of TCS AH. If not stated otherwise the given episode has been integrated/is scheduled for integration with TCS AH platform within the activities of the EPOS IP Project.

Table 6: Episode list of TCS AH

To be implemented			Implementation mode/ plan of technical implementation
Episodes (Data and Data Products)	Software	Applications	
BOBREK MINE: local seismicity linked to longwall mining			Integrated/IS-EPOS
LGCD: regional seismicity and ground motion associating underground hard rock mining			Integrated/IS-EPOS
USCB: regional seismicity and ground motion associating underground coal mining			Integrated/IS-EPOS
GROSS SCHOENEBECK: geothermal energy production experiment			Integrated/IS-EPOS
CZORSZTYN: Shallow water reservoir			Integrated/IS-EPOS
SONG TRANH: deep water reservoir			Integrated/IS-EPOS
GRONINGEN FIELD: conventional hydrocarbon production			Integrated/SHEER
LUBOCINO: Shale Gas			Integrated/SHEER
WYSIN: Shale Gas			Integrated/SHEER
PREESE HALL: Shale Gas			Integrated/SHEER, EPOS IP
THE GEYSERS: geothermal energy production			Integrated/SHEER
OKLAHOMA: conventional and unconventional hydrocarbon extraction and wastewater injection			Integrated/SHEER, EPOS IP

ASFORDBY: underground coal mining			Integration pending
LACQ FIELD: conventional hydrocarbon extraction			Integration pending
MONTEYNARD: water reservoir			Integration pending
PYHASALMI MINE: in situ underground laboratory			Integration pending
THE GEYSERS: Treated wastewater injection for geothermal power production			Integration pending
VAL D'AGRI FIELD: conventional hydrocarbon extraction			Integration pending
VAL D'AGRI: water reservoir			Integration pending
EMILIA ROMAGNA: cavone oil field			Scheduled for integration after PY2
GISOS-SOLVEY: underground solution mining			Scheduled for integration after PY2
NORTHWICH: underground salt extraction cavities			Scheduled for integration after PY2
SOULTZ-SOUS-FORETS: stimulation and production of geothermal energy			Scheduled for integration after PY2
THORESBY COLLIERY: underground coal mining			Scheduled for integration after PY2
GORDONSTONE COLLIERY: underground coal mining			Scheduled for integration after PY2[*]
RITTERSHOFFEN: stimulation and production of geothermal energy			Scheduled for integration after PY2
I2MINE: underground mining			Scheduled for integration after PY2
Virtual access to 2 superepisodes of MUSE (Multidisciplinary Upper Silesian Episodes)			Scheduled for integration/EPOS-PL

2.3. Metadata

Current activities concerning metadata within 14.3 package are related to integration of the existing TCS AH episode metadata with the metadata scheme of ICS-C.

The status of the metadata at the TCS AH level is the following:

- All relevant metadata are integrated at TCS level (currently 30k records)
  - Provides URLs for download data(files),
  - Some organization data missing (extension in progress),
  - Continues action for newly integrated episodes,
  - Simple, internal format built to support use cases on the platform and manage easily,
- Technical implementation for export/synchronization to JSON-LD in progress
  - DCAT with extensions in consideration
  - API:
    - traverse ALL metadata (paging),
    - list types,
    - follow relations,
    - update log (under consideration)

---

[*] Data providers are still negotiating level of production data availability for those three episodes with private companies.

The TCS AH metadata integration with ICS-C is scheduled for Q1 and Q2 of 2017. Metadata will be integrated with ICS-C scheme at various levels both, the episode level as well as particular episode building blocks level. Such approach is appropriate for EPOS users taking in to account their various needs.

3. TCS roadmap

All the Anthropogenic Hazards DDSSs planned for the EPOS IP project have been included in the priority list section. The sections in chapter: Priority List of DDSS, provide details of the roadmap for the DDSS integration with the TCS AH platform. To date, there are no delays in the integration plan implementation. All the harmonisation and TCS AH – ICS-C interoperability matters are worked out in close collaboration with EPOS WP6 and WP7.

4. Data Management Plan (DMP)

Each TCS should have a DMP for their own sustainability during the operational phase of EPOS (EPOS-OP).
It would be essential that each DDSS element defines its own DMP discussing the following topics: i) Data Access Policy, ii) Data storage and maintenance responsibility, iii) Data curation responsibility, iv) Data management and governance structure, v) Financial commitment securing the operational costs

TCS AH has distinguished several classes of user and determined its rights. This is the following: USERS:

Individual Researcher (R). TCS AH integrates research infrastructure to facilitate research hence R- class users are principal users of TCS AH. R is authorized by TCS AH platform admin upon his/her own request. This class membership is resolved on the basis of applicant's affiliation.

Internal Project Participant (PP). TCS AH is going to be an infrastructural pillar, around which AH community integrates and organizes to carry on collaborative research. TCS AH will support collaborative functions including call for project issued by a TCS AH user, declaration of interest in participation from other users, tools facilitation structuring such an internal project, common workspace for the project, which is shared by the project members, a possibility to upload/download data and codes to the common workspace, tools for oral communication of project participants etc. The most often PP is a researcher and then has also all the powers of R. If not, the powers of PP against the resources of TCS AH platform arise from belonging of the user to a specific users' class.

External Project Participant (EPP). A member of a research project, whose coordinator agreed with TCS AH management special conditions concerning the use of TCS AH infrastructural resources for the project purposes. EPP, registered and authorized upon request of the project coordinator, obtains access to restricted resources of the project, located on TCS AH platform. The most often EPP is a researcher and then has also all the powers of R. If not, the powers of EPP against the other resources of TCS AH platform arise from belonging to a class of users.

Representative of Industry Partner (IP). The TCS AH - industry partnership concept responds to the needs to raise contacts between science and industry to the level, at which the two parties

mutually impact one another. In the framework of this partnership, Industry Partner makes a part of its infrastructures, either hard (instruments, monitoring networks) or soft (data, including relevant operational), available for research through TCS AH. It is worth to note that research does not need the most recent data, which can be sensitive for Industry Partner. The comprehensiveness of the studied case description with data is that, which is essential. TCS AH management, from its side, implements the mechanisms, which prevent violation of rights of Industry Partners, including intentional or unintentional abuse of information that can be related to the Partner. Furthermore, Industry Partners provide advice on the potential usefulness and practical applicability of the solutions meant to be used in practice. In return, Industry Partners obtain possibility to convey to research community wider problems of their interest, and the TCS AH admin, through the internal communication lines of TCS AH platform, encourages researchers to focus their attention on these problems. Industry Partners have priority in the implementation of services developed be the research community grouped around TCS AH. Tests of these services in operational conditions of industry will positively influence their further development. The TCS AH – industry partnership has the form of a signed agreement between the two parties IP is authorized by TCS AH platform admin upon request of the management of the Industry Partner.

Representative of Institutional User (IU). Various interactions between TCS AH and either academic of non-academic organizations/institutions are envisioned. In every case its specific form, appropriate for the Institutional User's needs and TCS AH strategy will be determined in a signed agreement between the two parties. The agreement will define the access extent to the resources of TCS AH. IU is authorized by TCS AH platform admin upon request of the management of the Institutional User.

Trainer/Trainee (T). TCS AH is a perfect tool for training and educational purposes. Institutions, which will decide to use TCS AH for training/education, will agree the use of TCS AH resources for these functions with TCS AH management. T is authorized by TCS AH platform admin upon this agreement.

Public (P). A registered but not authorized visitor of the IS-EPOS platform.

INFRASTRUCTURE:
EPISODES:

E-A – Fully accessible for all users.
E-P – Project episodes. The episodes are visible for all users but open only for Project Participants (PP) throughout a prescribed time period. After this period the episodes become E-A.
E-B – Embargoed. This are episodes sensitive for the episode provider. Upon the episode provider's request the TCS AH platform management restricts they use throughout a prescribed time period. The episodes are visible for all users but the embargo conditions define their accessibility. After the embargo period the episodes become E-A.

SOFTWARE:

S-A – Services to be used by all.
S-B – Services, whose proper use require a specialized background of the user. Tools, whose use could be commercialized.
S-C – Codes based on newly proposed methods and algorithms for tests and developments.

DOCUMENTS: No classes are distinguished
ADDITIONAL FUNCTIONALITIES:

Workspace Area – The user obtains a quota of workspace, in which he/she can store data, process data, store partial results and final results of processing. The user's workspace is not visible for other users.

Workspace Sharing – The user can share a part of his/her workspace with other users. Collaborative Functions – Functionalities facilitating setting and running either internal or external project.

Data Upload – The user can temporarily upload data and process it using the TCS AH services. This data is uploaded to the user's workspace and hence cannot be used by other users unless the workspace sharing facility is enabled.

Episode Download – The user can download the integrated data. This functionality is envisaged and will be implemented after the detailed IPR policy is worked out.

Results Download – A possibility to download results of processing: numbers, text, tables, graphs. The exception is the low level services' results download, which undergo the rules of episode download.

Software Download – An option enabling the user to download open licensed codes.

Document Upload – The user can upload documents, which are reviewed by the editor of the document repository. Upon editor's acceptance the documents are made public in the repository.

USER ROLES BREAKDOWN

		R	PP	EPP	IP	IU	T	P
Episode Access	E-A	✓	✓	✓	✓	✓	✓	✓
	E-P			✓				
	E-B	✓ [1]	✓ [1]	✓ [1]	✓ [2]			
Workspace Area		✓	✓	✓	✓	✓	✓	
Workspace Sharing			✓	✓			✓	
Collaborative Functions		✓	✓	✓				
Data Upload		✓	✓	✓	✓	✓ [4]		
Episode Download		✓	✓	✓				
Results Download		✓	✓	✓	✓	✓	✓	
Software Access	S-A	✓	✓	✓	✓	✓	✓	✓
	S-B	✓	✓	✓	✓ [3]	✓ [4]	✓ [5]	
	S-C	✓	✓	✓				
Software Upload		✓	✓	✓				
Software Download		✓	✓	✓	✓			
Document Access		✓	✓	✓	✓	✓	✓	✓
Document Upload		✓	✓	✓				

[1] Based on the embargo definition access may be limited to certain user groups.
[2] If embargoed by this IP.
[3] According to conditions of the partnership with this IP.
[4] According to conditions of the agreement with this IU.
[5] According to conditions of the course/workshop.

Governance structure and financial commitments are described respectively in two deliverables:
D14.1 Report on Legal and Governance Framework for TCS Implementation of WP14
D14.2 Report on financial work and sustainability for TCS implementation of WP14

5. Use cases

Even if use cases are excellent for showing the possibilities of tasks which can be performed on AH episodes via TCS AH platform they also limit research options. Nevertheless, a simple use case being currently under development for TCS AH platform is provided below. We would like to stress however, that TCS AH will look for options to leave as much research freedom to the users, as possible, by finding possibilities of providing an "active" platform in contrast to "static" one based on already predefined use cases.

Use case name/topic: anthropogenic mine hazard
Use case domain This use case is: Anthropogenic Hazards   o   multidisciplinary, namely focusing on the disciplines of: seismicity, geology, GNSS data, other gathered within a mining episode.
Use case description in this section the use cases will be outlined. This section may require iterative refinements. *As a user, I want to obtain set of parameters characterising stationary hazard and their time dependence.* …

Actors involved in the use case A list of the actors who communicate with this use case.
 	
    •   user
    •   user with special privileges to access time embargoed data

Priority How important is this Use Case to the TCS? (considering the degree of use: unavoidable/ frequent/occasionally/rare)   o   High
Pre-conditions A list of conditions that must be true before the Use Case starts (e.g. user must have logged in) CIBIS episode data centres have to be interconnected with TCS. Required software applications have to be integrated with the IT platform and connected to computing resources.

Flow of events – user view
1. Data selection & preparation
 	
a. Data selection (AH episode) which user will analyse. In case of catalogue, it ought to contain all of: the date, localisation, event magnitude and/or its energy
 	
b. Additional data choice like front development, mine borders, geology/tectonics of the given area
 	
c. Define the coordinates for a given data set (Cartesian or/and geographic, if there is any missing it has to exist the possibility to add it).
 	
d. determine the threshold of completeness
 	
e. Data visualisation by:
 	
    • Histograms (with possibility to define ranges for a given data set)
 	
    • Point/line/column plots for selected data [localisation, mine borders, front development etc.].
 	
    • data sorting and selection via different criteria
 	
f. determination of the probability of the seismic zones with
 	
    A. the method based on the unlimited Gutenberg-Richter model,
 	
    B. the method based on the upper-bounded Gutenberg-Richter model
 	
    C. the unbounded non-parametric distribution estimation

D. the upper-bounded non-parametric distribution estimation g. Zone parameterisation and probability characteristic check h. Additionally, user can request zone gravity centre and other zone related parameters. **2.** A) Determination of Stationary hazard parameters for a given zone Data resulting from computations shall be visualised on a map **2.** B) Determination of the time depended hazard parameters For each zone user can determine the hazard parameters time characteristics during given zone activity. This include a time window definition for a given parameter and final visualisation.
System workflow - system view • Will be definitely of use for management/extraction/ of the data and for determination hazard parameter sets.
Post-conditions visualisation services for data presentation
Extension Points depends heavily on user needs.
« Used » Use Cases
Other Requirements This can include non-functional requirements related to the Use Case. …
(to be filled in by WP7) After the interview: create class and sequence diagram for each use case. Class diagram and sequence diagram.

6. CONCLUSIONS

The service provided by WP14: is the Thematic Core Service Anthropogenic Hazards platform. The platform includes generic tools required by the AH community as well as episodes (gathering multidisciplinary data and data products from many sources) and software applications allowing analysis of those and speeding up the research conduction. The activities related to TCS AH platform include both: the integration of the resources (aps, episodes) and interoperability with ICS-C central hub together with unification of data formats according to EPOS standards.
The level of integration advancement is in line with WP14 development plan.