Ref. Ares(2018)4999926 - 29/09/2018

VRE4EIC

A Europe-wide Interoperable Virtual Research Environment to Empower Multidisciplinary Research Communities and Accelerate Innovation and Collaboration

Deliverable D2.6

Use-case report – Second Version

Document version: 2.0 - 29/09/2018

VRE4EIC DELIVERABLE

Name, title and organisation of the scientific representative of the project's coordinator:

Mr Peter Kunz t: +33 4 92 38 50 10 f: +33 4 92 38 78 22 e: peter.kunz@ercim.eu

GEIE ERCIM, 2004, route des Lucioles, Sophia Antipolis, F-06410 Biot, France

Project website address: http://www.vre4eic.eu/

Project	
Grant Agreement number	676247
Project acronym:	VRE4EIC
Project title:	A Europe-wide Interoperable Virtual Research Environment to Empower Multidisciplinary Research Communities and Accelerate Innovation and Collaboration
Funding Scheme:	Research & Innovation Action (RIA)
Date of latest version of DoW against which the assessment will be made:	31 May 2017 Amended Grant Agreement through amendment n°AMD-676247-8
Document	
Period covered:	M1-36
Deliverable number:	D2.6
Deliverable title	Use-case report – Second version
Contractual Date of Delivery:	30/09/18
Actual Date of Delivery:	29/09/18
Editor (s):	EuroCRIS
Author (s):	Muckensturm Marc, Dragan Ivanovic, Laurent Remy (euroCRIS), Zhiming Zhao (UvA), Daniele Bailo (INGV), Anneke Zuiderwijk (TU Delft)
Reviewer (s):	Carlo Meghini (CNR), Jacco Van Ossenbruggen (CWI)
Participant(s):	TUDelft, INGV, UvA
Work package no.:	2
Work package title:	Requirements, use cases and usability, and evaluation
Work package leader:	Anneke Zuiderwijk (TU Delft)
Distribution:	PU
Version/Revision:	2.0
Draft/Final:	Final
Total number of pages (including cover):	64

What is VRE4EIC?

VRE4EIC develops a reference architecture and software components for VREs (Virtual Research Environments). This e-VRE bridges across existing e-RIs (e-Research Infrastructures) such as EPOS and ENVRI+, both represented in the project, themselves supported by e-Is (e-Infrastructures) such as GEANT, EUDAT, PRACE, EGI, OpenAIRE. The e-VRE provides a comfortable homogeneous interface for users by virtualising access to the heterogeneous datasets, software services, resources of the e-RIs and also provides collaboration/communication facilities for users to improve research communication. Finally, it provides access to research management /administrative facilities so that the end-user has a complete research environment.

Disclaimer

VRE4EIC has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 676247.

Table of Contents

Contents

1 Introduction *

Following the first requirements collection (see D2.1), use-cases are elaborated to express the possible users' interactions with the system.

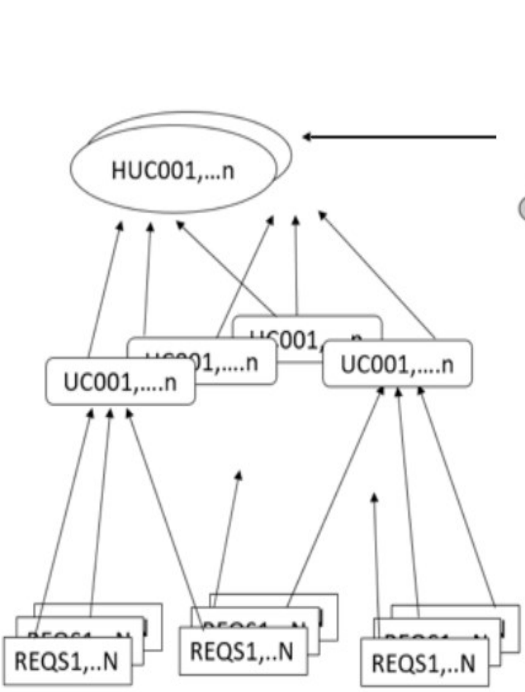
Use-cases can be, and are, expressed at several levels of granularity:

- Step 1: (Low level) use-cases are built by assembling requirements in a coherent sequence;
- Step 2: High-Level use-cases are described re-using the (low level) use-cases;
- Step 3: "Visionary" use-cases are independent of the requirements, represent several scientific domains and reflect the advanced research goals of the project.

In the first version of this deliverable (D2.3[1]), 59 use-cases and 19 high-level use-cases have been identified and described. 26 "additional" requirements have also been identified in order to fill gaps in the use-cases flow. Several steps were still to be accomplished, namely:

1. The relation between the visionary use-cases and the High-level use-cases or directly with the Use-cases should be analysed; EPOS and ENVRIplus use-cases will be compared to VRE4EIC High-Level Use-cases or Use-Cases
2. Use-cases should be used to validate the architecture design
3. Use-cases and additional requirements should be brought to the potential users to get their feedbacks

In this second version, 13 visionary use-cases have been identified and mapped to the High-level use-cases. The following figure represents the relations between requirements and use-cases, as stated in D2.3, including the visionary use-cases.

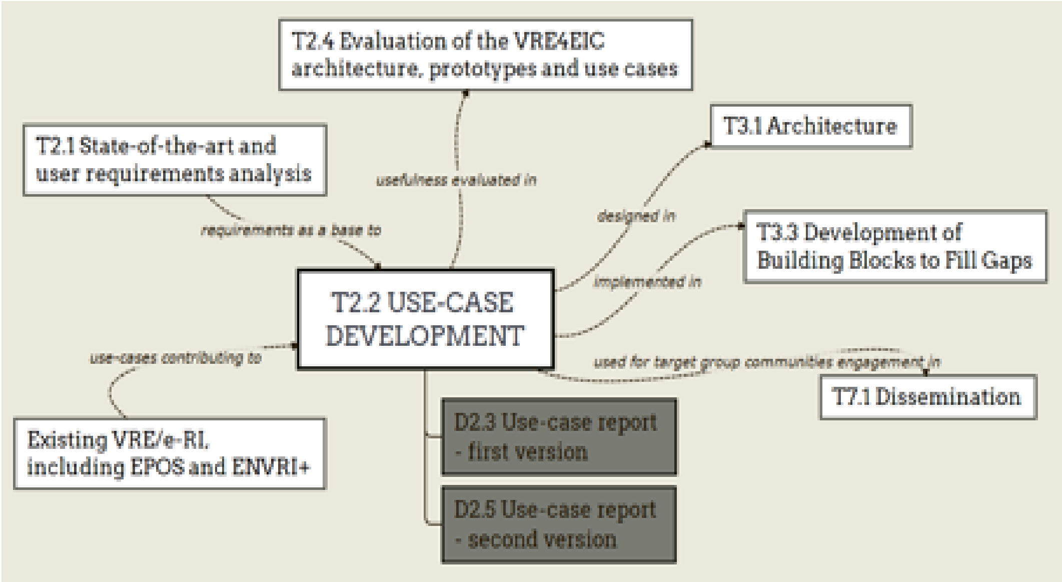


[1] https://www.vre4eic.eu/images/Public_deliverables/D2.3_Use_case_report-first_version.pdf

Additionally, this second version also presents the validation of the architecture design regarding the use-cases, as well as relations that have been identified to use-cases developed in the context of EPOS and ENVRIplus projects.

Note: as this document is an update of the first version of the deliverable, some of its parts have been kept from D2.3; the updated or new chapters have been marked with an asterix *.

2 Relationship between the tasks



The T2.2 Use-case development task follows the T2.1 *State-of-the-art and user requirements analysis* task. It relies on the requirements list obtained by T2.1 to design the use-cases. It also uses the experience from the EPOS and ENVRIplus projects.

The designed use-cases based on the requirements gathered in D2.1 are also linked to the components designed in T3.1 Architecture on the basis of D2.1, and partially implemented in T3.3 *Development of Building Blocks to Fill Gaps.*

The use-cases have also been used to engage target group communities within T7.1 *Dissemination*.

The T2.4 *Evaluation of the VRE4EIC architecture, prototype and use-cases* task collected feedbacks on the usefulness of the use-cases and on to which extent the use-cases can be implemented.

3 Semi structured documentation of use cases *

The use-case documentation template has been described in the first version of this deliverable (D2.3). This second version will focus on visionary use-cases, which template is inspired from the use-case template, and is described below.

For each visionary use-case, a document describes it using the following sections:

Visionary Use Case Title

<The title of the use-case>

Disciplines

<A list of scientific disciplines involved in the use-case.>

Short case explanation

<A text explaining the purpose of the visionary use-case.>

Visionary Use-case description

<A formal description of the use-case: “As a User, I want to…”>

Actors

<A list of the Actors who communicate with this Use Case>

Priority

<How important is this Use Case to the project>

Pre-Conditions

<A list of conditions that must be true before the Use Case starts>

Flow of events - user view

<This could be in text, in an Activity diagram, in a Sequence diagram, or any or all of them. This flow is user-centered.>

System workflow - system view

<This could be in text, in an Activity diagram, in a Sequence diagram, or any or all of them. This flow is system-centered.>

Post-Conditions

<A list of conditions that must be true when the Use Case ends, no matter which Scenario is executed.>

4 Requirements

4.1 Requirements from D2.1

The collected requirements (D2.1) have been classified in the following categories:

1. Data identification and citation (prefix: IRQ)
2. Data curation (prefix: CRQ)
3. Data cataloguing (prefix: CLRQ)
4. Data processing (prefix: PRQ)
5. Data optimisation (prefix: ORQ)
6. Data provenance (prefix: PVRQ)
7. Collaboration, training and support (prefix: CTRQ)
8. Privacy, security, trust and legal requirements (prefix: SRQ)
9. FURPS+ and ISO 25010:2011 (non-functional requirements, prefix: NRQ)

Note: the FURPS+ acronym, devised by Robert Grady of HP, refers to the non-functional requirements categories named Functionality (Generality of Feature Set, reusability, security), Usability, Reliability, Performance, Supportability, and the constraints (+) on design, implementation, interface and physical properties, of the system.

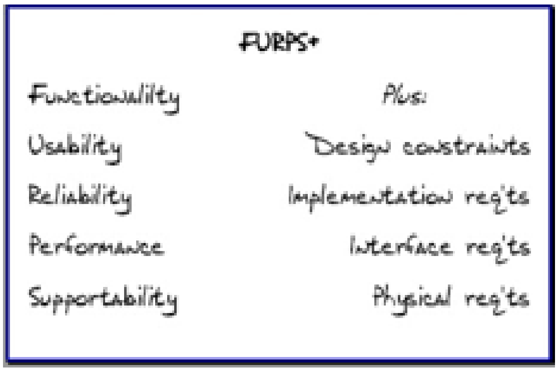


Figure 4.1.1 FURPS+ flash card (Ottinger and Langr, 2011)

The set of requirements has first been analysed category by category to identify use-cases by aggregating requirements from a category. Then, the use-cases have been completed by referencing requirements or other use-cases from any category.

Finally, the high-level use-cases are defined across categories.

4.2 Additional information on requirements

A survey and e-RI characterizations are 2 means used in addition to the initial interviews that allowed to gather the initial requirements.

The survey brings information about the level of expectation from (potential) users related to the requirements, thus allowing to priorities the requirements.

The e-RI characterizations bring light on the identified requirements that are or will be implemented at RI level. With this information, a clearer picture can be obtained of the part of requirements and thus use-cases, that can be executed at RI level and orchestrated at VRE level, versus the one that are executed at VRE level.

5 Use-cases

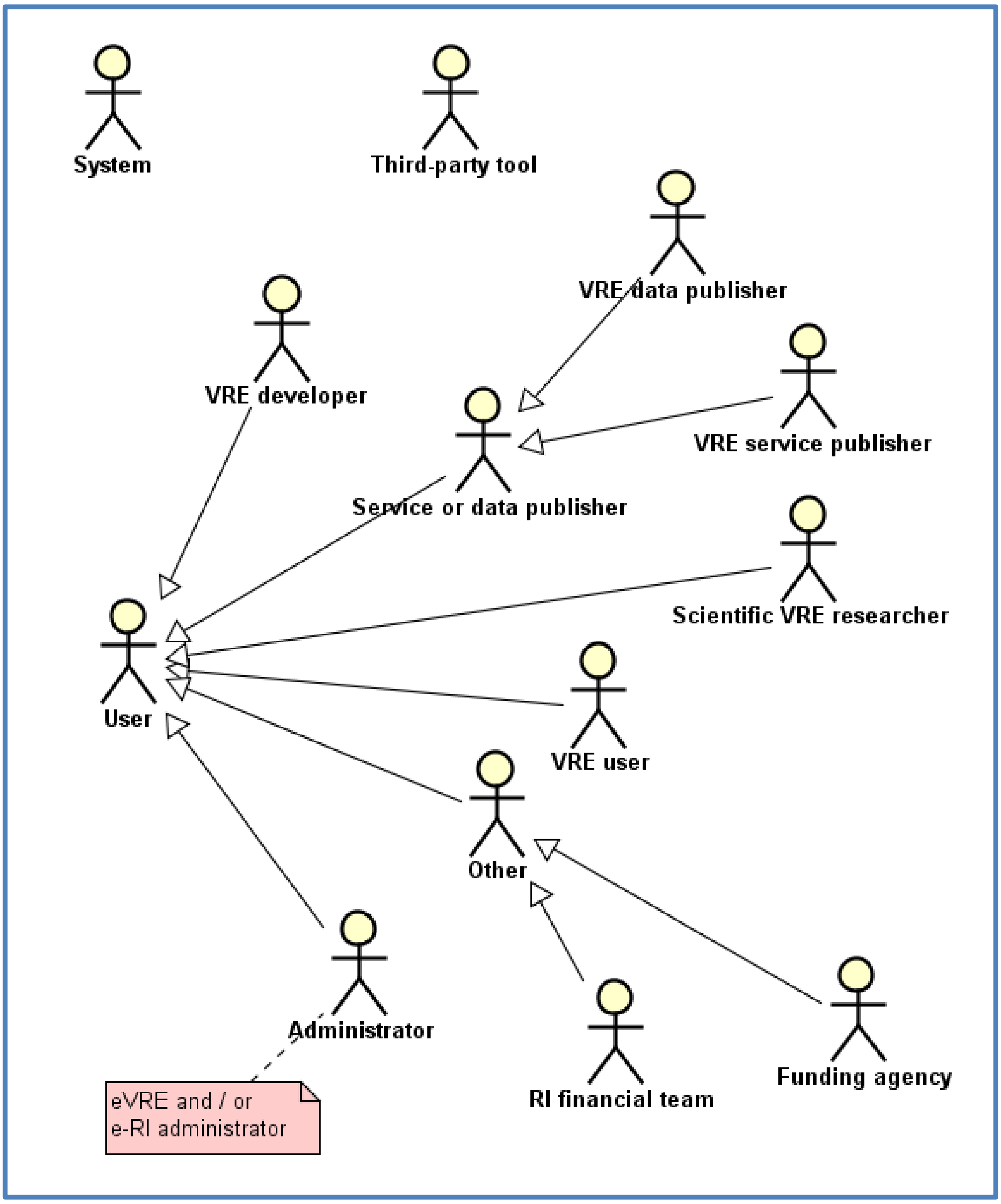
In this section, requirements from D2.1 are orchestrated into use-cases, and the actors executing these use-cases are given.

Requirements, and therefore use-cases, have been defined at several levels of details or granularity. To make these levels explicit, some use-cases "include" other smaller use-cases (to avoid repeating them).

Another relationship is the "extends" one. A use-case extending another one has additional steps.

5.1 Actors

Here are the actors fulfilling the roles described by the use-cases:



They can be described as such:

- User: global actor representing any user accessing the eVRE (according to its definition, "VRE users" only concern people that want to access research data)
    - VRE developer: including commercial (large IT companies, SMEs, entrepreneurs) and non-commercial (universities, not-for-profit organisations, foundations, VRE related projects) developers
    - VRE user: including academic and governmental researchers, research managers, educators, students, innovators, entrepreneurs and the interested citizens
    - Scientific VRE researcher: including academics who conduct research on VREs, for instance on VRE components and VRE communities
    - Service or data publisher: this actor represents any user that register services in the eVRE, or that publish some data.
        - VRE data publisher: i.e. publishers who wish their data to be available to VRE users, including research institutions and archives, universities, governmental organisations, various researchers and other data publishers
        - VRE service publisher: i.e. publishers of services that can be triggered from the VRE environment
    - Other: at the same time, we envision other potential target groups, such as journalists, educators and students, although these groups are not key to the project
        - Funding agency: funding agencies that publish calls for research fundings
        - RI financial team: the financial team of a specific RI
    - Administrator: a specific user that has specific rights to manage users in the platform
- System: this actor represents the eVRE platform itself, for automatic behaviors
- Third-party tool: this actor represents tools not included in the eVRE platform

5.2 Use-cases per category

By orchestrating the requirements under different categories, we have defined distinct corresponding use-cases.

For more details about the specific categories and following use-cases, please refer to the deliverable D2.3 Use-case report - First version[2].

[2] D2.3 Use-case report – first version: https://www.vre4eic.eu/images/Public_deliverables/D2.3_Use_case_report-first_version.pdf

5.3 Summary of “additional requirements” identified

Additional Requirements

- ARQ001 Define things to be notified about
- ARQ002 Access from email link
- ARQ003 Merge 2 datasets
- ARQ004 Select data across several sources
- ARQ005 Transform data
- ARQ006 view workflows
- ARQ007 view workflow instanciations
- ARQ008 manage curation rules
- ARQ009 Standardised formats
- ARQ010 Datasets comparison
- ARQ011 Metadata conversion
- ARQ012 Register projects
- ARQ013 Assign project to team
- ARQ014 Publish funding agency information
- ARQ015 Change user profile
- ARQ016 Remove user
- ARQ017 Search funding proposal
- ARQ018 View funding proposal
- ARQ019 Download funding proposal
- ARQ020 Publish funding calls
- ARQ021 Search funding calls
- ARQ022 View funding call
- ARQ023 manage my expertise
- ARQ024 Create a dataset
- ARQ025 Select data from a dataset
- ARQ026 Metadata comparison

6 High-Level Use-cases *

All high-level use-cases (HUC) have already been described in the first version of this document. Please refer to D2.3 Use-case report - First version[3].

The list of High-level Use-cases is the following:

1. HUC001 Access services and data from eVRE
2. HUC002 Annotate data
3. HUC003 Compare datasets
4. HUC004 Create dataset
5. HUC005 Create dataset from instrument
6. HUC006 Manage data
7. HUC007 Manage funding calls
8. HUC008 Manage instrument
9. HUC009 Manage research infrastructure
10. HUC010 Manage research projects
11. HUC011 Process data
12. HUC012 Publish dataset
13. HUC013 Transform data
14. HUC014 Manage personal profile
15. HUC015 Manage services offered in eVRE
16. HUC016 Manage users
17. HUC017 Query data
18. HUC018 Communicate
19. HUC019 Get support

[3] https://www.vre4eic.eu/images/Public_deliverables/D2.3_Use_case_report-first_version.pdf

7 Visionary Use-cases *

7.1 Introduction

By starting with an existing realistic set of use cases and high level use cases previously discovered, we explored ideas from which to deduce and then develop more hypothetical or visionary use-cases.

A visionary use-case consists of an exploration of data that would be needed in a specific scenario, with consideration to the requirements for a future use of multidisciplinary research data.

We were able to characterize 13 visionary use cases.

7.2 Gathering methodology

The first set of visionary use cases were derived using a brainstorm session by the VRE4EIC project partners. Each partner thought about what is currently possible with VREs and what would ideally be possible with VREs. Visionary use cases were developed considering an ideal situation. The visionary use cases covered different scientific disciplines and different activities. From this first gathering round, 10 visionary use cases were derived.

A second set of visionary use cases was collected through a workshop with TU Delft students. A one-hour workshop was organized at the one lecture of Bachelor course TB321: Information Communication Governance at TU Delft, The Netherlands, on 22nd March, 2018. TU Delft prepared three questions for this workshop. 13 Bachelor students attended the workshop and they were divided into three groups for discussion. Three additional visionary use cases were derived from this workshop.

7.3 Description of visionary use cases

In this section, the 13 visionary use-cases are shortly described. The full description can be found in annex A at the end of this report.

The visionary use-cases are descriptions of detailed interactions between actors and the system to be built. They capture <u>who</u> (actors) does <u>what</u> (interactions) with the system, for <u>what purpose</u> (goal). Regarding the actors, please note that all the actors identified below are of the type “VRE user” in the actors identified for our eVRE in chapter 5.

VUC 1: Investigating historical interest based on current events

As a <junior digital humanities researcher> I want to <analyse usage logs from different Research Infrastructure (RIs)> in order to <discover whether there is a renewed interest in the Great Depression in Europe between 1929-1940 since the financial crisis has hit the EU>.

As a <senior digital humanities researcher> I want to <analyse usage logs of the Dutch REs> in order to <discover where interest is focused on with respect to the Netherlands and where this interest is originating from>.

VUC 2: Investigating capacity of urban infrastructure in case of evacuation

As a <governmental researcher> I want to be able to <view and visualize on an Open Street Maps map, a dataset regarding the frequency and impact of volcano explosions and plot in the same map a dataset regarding the urban/transport infrastructure in a certain area and see citizen counts related to nearby cities on the same map> so that I can <gain an understanding of capacity and level of sufficiency of the available urban/transport infrastructure and housing in nearby cities in case of an emergency>.

VUC 3: Evolution of electric charging

As a <(governmental) researcher> I want to be able to <view and compare the availability of charging points in an EU city/country A and the number of electric vehicle (EV) in A> so that I can <gain an understanding of the level of sufficiency of the number of charging points in A and the ability to increase the amount of charging points>.

VUC 4: Relation between age of first pregnancy and level of education followed

As <a researcher> I want to be able to <view and compare the age of first pregnancy under women and the level of education in a EU country> to <gain insights in the relation between age of first pregnancy and level of education>.

VUC 5: Social consequences of increased debts of EU countries

As <a researcher> I want to be able to <view and compare the GDP of country X and Y and their unemployment rate respectively> to <gain insights in the relation between the debts with respect to their GDP and unemployment rate and the criminality rates>.

VUC 6: Shipwreck archaeology

6.1 As an <earth science researcher> I want to <collect remote sensing data from satellites, such as the images of earth surface> to <store in the e-VRE>.

6.2(a) As a <marine environment researcher> I want to <download the historical marine tide data> to <test my tide flow model>.

6.2(b) As a <marine engineering researcher> I want to <use the images captured by satellite and tide and current simulation data> to <design a modelling method to detect the location of the shipwrecks close to coastal line>.

6.3 As an < archaeologist> I want to <use the location data of shipwrecks > to <conduct underwater archaeological excavation at a more accurate location>.

VUC 7: Prediction of transport delay

As a <(transportation) researcher> I want to be able to <view and compare travel time and weather conditions for some region> so that I can <gain insights in the relation between travel time and weather condition> in order to create a model which could better predict transport delay and better organize connected transports.

VUC 8: Choosing travel destination for tourists with allergic diseases

As a <(healthcare) researcher> I want to be able to <view and compare the environmental data and allergic dataset of a region> to <gain insights in the relation between the time of the year, weather and the allergy rates for the region> in order to create a warning/recommendation system for tourists with allergic diseases before choosing travel destination in the certain period of the year.

VUC 9: Impact of lifestyle on health

As a <health researcher>, I want to study different <people's lifestyles on smoking, drinking and diet> with the aim to <find some correlations with certain chronic disease> in order to publish these relations, to prevent and create awareness of risks on certain lifestyles.

VUC 10: Urban development in Netherlands

As a <city planner> I would like to be able to <access data from the public transport cards in the Netherlands> so that I can gain an <understanding of how people move through the city> in order to optimize the public transport in the city.

VUC 11: A.I. to diagnose mental health issues

As a <neuroscientist> I want to <test people with mental health issues> so that I can <gain insight in patterns shown by people with mental health issues> in order to <develop an AI to diagnose and help these people>.

VUC 12: Locate DDoS attacks

As a <researcher> I want to <be able to analyse the traffic flow of DDoS amplification attacks> so that I can <gain insight in the security practices in different regions> in order to develop security governance practices in weak areas.

VUC 13: Seismic activity and global warming

As a <researcher> I want to be able to <map and deduced that increasing seismicity will indicate increasing geothermal flux and to compare the earthquake activities with the temperature since the past year> so that I can <gain an understanding of how geothermal heat activities on earth contributes to rise the average world temperatures> in order to warn the climate community to explore the impacts of geothermal flux.

7.4 Mapping of visionary use cases

In this section, we check that visionary use cases are covered and can be expressed as a combination of existing use cases. For this task, the visionary use cases (VUCs) previously described have been mapped to high level use cases (HUCs).

In the following table, VUCs are presented in columns and HUCs by rows. When a VUC includes at least part of a HUC, a check is placed at the intersection of the corresponding column and row. Thus we can directly know if a HUC is not included in any VUC (row empty) or if a VUC is not realized by any HUC (column empty).

VUC	1	2	3	4	5	6			7	8	9	10	11	12	13
						6.1	6.2	6.3							
HUC001 Access services and data from eVRE	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
HUC002 Annotate data	X					X	X				X		X		
HUC003 Compare datasets	X	X		X	X				X	X	X	X	X		X
HUC004 Create dataset	X	X				X	X			X	X		X	X	X
HUC005 Create dataset from instrument					X	X	X			X	X		X		
HUC006 Manage data	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
HUC007 Manage funding calls															

HUC008 Manage instrument					X				X	X					
HUC009 Manage research infrastructure															
HUC010 Manage research projects	X	X									X				
HUC011 Process data	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
HUC012 Publish dataset			X			X	X				X	X	X	X	X
HUC013 Transform data	X	X		X	X				X	X	X	X	X	X	X
HUC014 Manage personal profile															
HUC015 Manage services offered in eVRE	X	X	X			X	X		X	X	X	X	X	X	X
HUC016 Manage users	X														
HUC017 Query data	X	X	X	X	X	X	X	X						X	X
HUC018 Communicate	X							X			X				
HUC019 Get support															

For this mapping, all the visionary use cases are supported by high-level use-cases developed, and most of the high-level use-cases are used by at least one visionary use-case.

Four high-level use-cases are not used by visionary use-cases:

- HUC007 Manage funding calls
- HUC009 Manage research infrastructure
- HUC014 Manage personal profile
- HUC019 Get support

These high-level use-cases are very specific, either to an actor, not represented in visionary use-cases, or to an activity that is linked to the platform management (support, profile management) or to the management of the research infrastructure. As the visionary use-cases were focused on the actor "VRE user" and the use of interdisciplinary data, these four HUCs are out of scope in this specific context.

8 Alignment with EPOS/ENVRIplus use cases *

8.1 EPOS

High level use cases and included Use Cases that are relevant for EPOS, are reported in the table below. VRE4EIC covers almost all use cases envisaged by EPOS except a few which are described by means of the diagrams below.

HUC001	Access services and data from eVRE
UC028	Be identified on VRE
UC033	Access the eVRE
UC039	Access services and data from eVRE through API
HUC017	**Query data**
UC016	Access a curated list of resources references
UC020	Search and retrieve datasets
UC022	Define a set of resources for an action (search, access)
UC031	Provenance of data and data usage information
UC050	Log Access Data
UC051	Usage data rules
UC052	Query data from a source

Table 8.1.1: High level use cases and included Use Cases that are relevant for EPOS

Diagram 8.1.2 represents the VRE4EIC High Level Use-case HUC017 where the differences have been found. UC020 and UC016 are the original use-cases that describe HUC017. Additional EPOS functionalities have been added to this high-level use-case to show the gaps between both cases.
A subset of the EPOS metadata was ingested and hence integrated into the VRE4EIC metadata service, so it is possible to Search and retrieve a dataset by using the VRE4EIC Graphic User Interface. In the first version of the use-case report D2.3, it was not specifically defined that a user should be able to search using a bounding box. Details about dataset visualization and interface drawing were also not represented in the requirements. The ingestion of EPOS metadata however changed the vision, and these features have been developed in the prototype. So the “Bounding Box Search”, “Draw interface”, “show dataset detail” functions are available. The same applies for “View Dataset (on) map” that is not covered currently in the VRE4EIC prototype, due to the additional visualization modules required. Also, this last functionality is very important according to the EPOS community, but might be an extension of less importance for other communities.

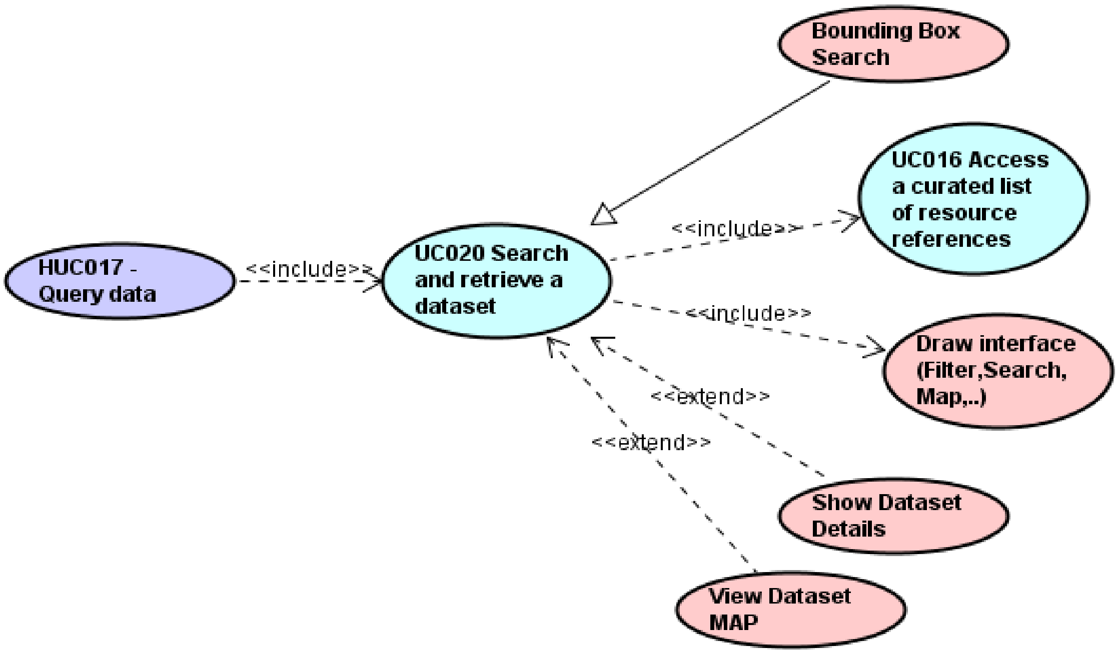
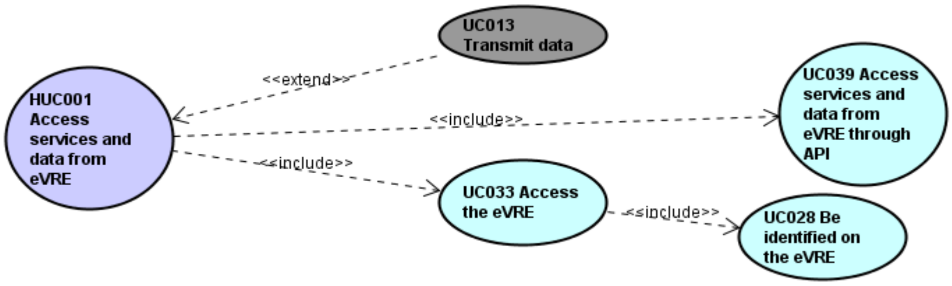


***Diagram 8.1.2**: Purple and cyan circles respectively represent High Level Use-cases and Use-cases in VRE4EIC; red circles represent EPOS functionalities to be included in a VRE (hence VRE4EIC)*

Similarly to the previous diagram, the following diagrams represent High Level Use cases defined by VRE4EIC with linked use-cases that are supposed to be covered by an enhanced VRE, but are not -or not yet- covered by EPOS. A short description and status of each of the grey use cases follows:

1. *UC013*: this use case requires to implement a "transmit data" functionality that should be available on the remote platform. This is not yet implemented in the EPOS platform but it is planned as a future development; Note: the transmission of data in UC013 involves all kind of transmission: from the eVRE platform or the underlying e-infrastructure to the computer of the user as well as between various e-infrastructures
1. *UC051*: this use case is about providing an authorization schema to any of the resources (e.g. dataset) in EPOS. Currently it is a work in progress in EPOS, and will require a long time due to the complexity of trans-national authorisation schema standards development;
2. *UC050*: this use case requires to log access data, i.e. information about access data including which data was accessed. It is currently a work in progress in EPOS;
3. *UC052*: this use case requires the ability to query data from a source (e.g. web-services) within the EPOS domain. It is currently developed, as the EPOS system architecture envisage a main node accessible to the public that integrates resources from RIs by means of (among others) web-services (i.e. data sources);
4. *UC031*: this use case requires to track provenance information of the accessed data. Those information are however many and of different kind: they span from data usage information to publication information, acknowledgement of contribution by other RIs, history of the data object and others. In EPOS, the provenance information are included up to a certain level (e.g. reference to source, user, license) but more advanced provenance information (like history of data object) is not available yet.



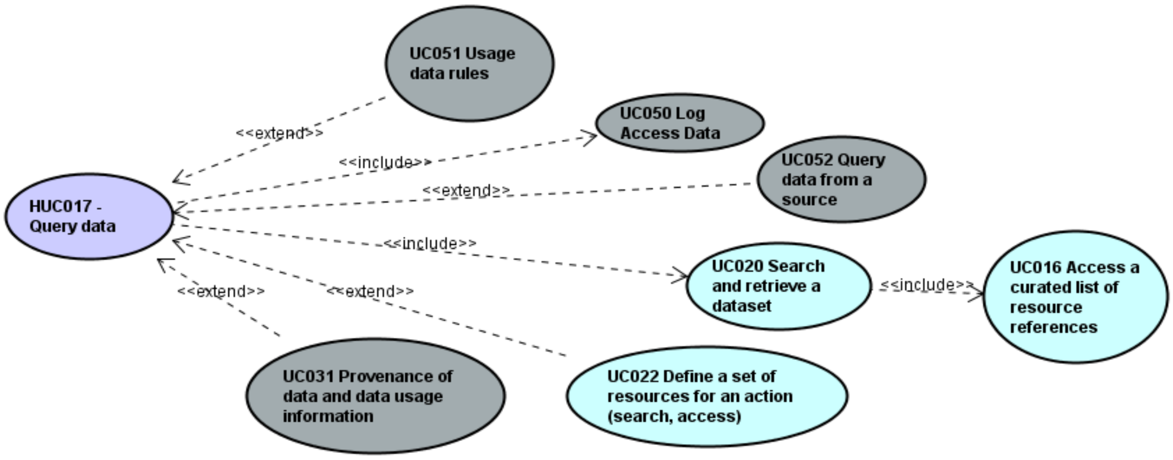


***Diagram 8.1.2**: Purple and cyan circles respectively represent High Level Use-cases and Use-cases in VRE4EIC; grey circles represent VRE functionalities that should be included in EPOS as a VRE*

In conclusion, EPOS and VRE4EIC share a global view on a VRE. Some variation can be found, but they only concern details of the implementation of some specific functions (logs, visualization for example).

8.2 ENVRIplus

The ENVRIPLUS project tackles common problems in the construction of e-RIs in the environmental and earth science domain by developing and evaluating reusable solutions. The use-cases identified in VRE4EIC and ENVRIPLUS, shown in Figure 8.2.1, are clearly complementary and well-aligned. The yellow boxes represent VRE4EIC High-Level Use-cases when the blue coloured blocks are specific use-cases from ENVRIPLUS.

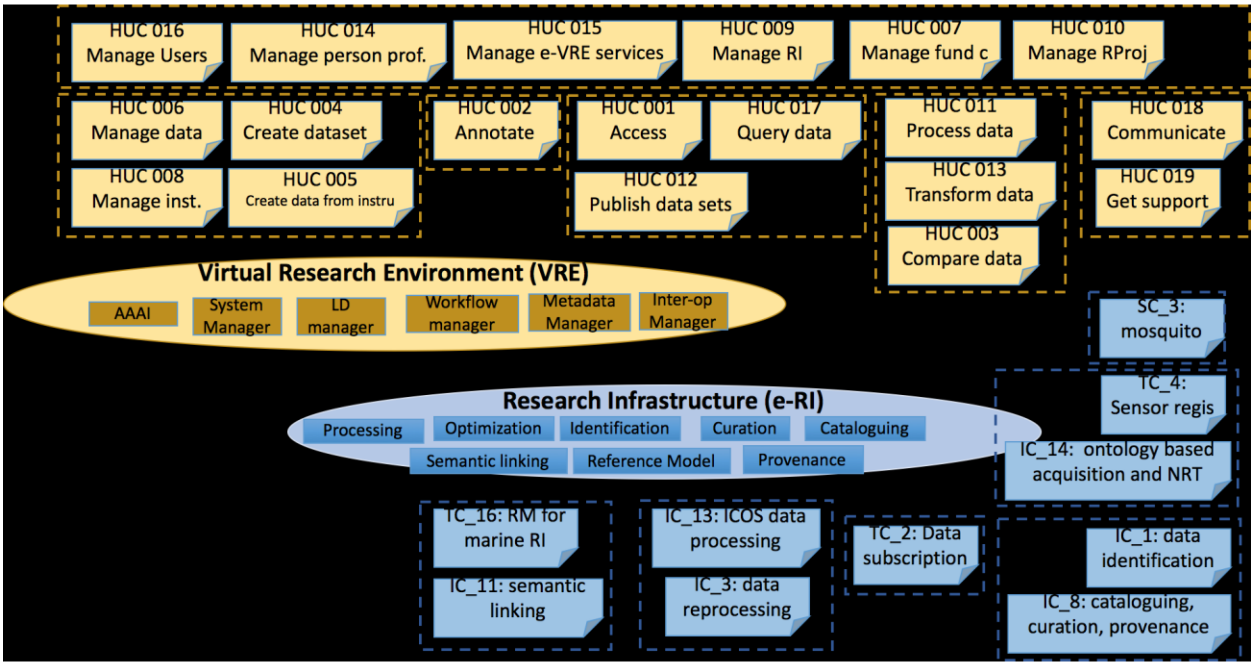


Figure 8.2.1: *Cases from VRE4EIC and ENVRIPLUS[4].*

To properly evaluate the quality of alignment between VRE4EIC and ENVRIPLUS use cases however, we need to ask the following two questions:

- Can ENVRIPLUS e-RIs support VRE4EIC use cases?
- Can VRE4EIC enhance ENVRIPLUS use cases?

8.2.1 ENVRIPLUS e-RIs for VRE4EIC use case

Table 8.2.1.1 provides analysis on how VRE4EIC use-cases are supported by ENVRIPLUS. Through the service portfolio developed by the ENVRIPLUS, we can see that the ENVRIPLUS services of data acquisition, curation, cataloguing, identification/citation, publishing and use, can well support VRE4EIC use-cases. The VRE resource management will be done at the high level, which requires no direct support from ENVRIPLUS e-RIs.

Use cases in VRE4EIC	Relevant services in ENVRI RM	Services to be developed in ENVRIPLUS theme2	How?

[4] The detailed use case lists from ENVRIPLUS can be found at: https://wiki.envri.eu/display/EC/Use+Cases

HUC9, 14 and 16	Partially	Cataloguing	Data, sensors, not users
HUC7,10 and 15	Not directly		Provided by VRE
HUC 4,5,6 and 8	Acquisition	No direct services to be developed in ENVRIPLUS. RIs use their existing acquisition services.	Sensor catalogues
HUC 2	Curation	Curation, Semantic linking	CERIF/OIL-E
HUC 1,12 and 17	Publishing	Identification, curation, cataloguing	DOI, CERIF/CKAN
HUC 3,11 and 13	Processing	Processing and optimization	D4Science, DRIP
HUC 18 and 19	Data use	Community support (not as ICT service)	E-Infrastructure solutions

***Table 8.2.1.1**: VRE4EIC high-level use-cases coverage by ENVRIPLUS*

8.2.2 VRE4EIC enhanced ENVRIPLUS use cases

The VRE4EIC solution can greatly enhance the ENVRIPLUS cases. The most important features include: workflow management, the interoperability manager, metadata, and the Linked Data manager. Table 8.2.2.1 shows the coverage of ENVRIPLUS use-cases by VRE4EIC high-level use-cases.

ENVRIPLUS use case	VRE4EIC architecture	VRE4EIC cases	How
SC_1: mosquito	Workflow, metadata, interoperability manager	**HUC 11, and 17**	
TC_4, sensor regis, TC_14, acquisition	Metadata, LD manager	**HUC 4,5,6 and 8**	Complement
TC_1, identification **TC_8, curation, cataloguing**	Metadata, LD manager	**HUC 2**	Cross e-RI linking, exploration and annotation
TC_2, subscription	Metadata, LD manager, interoperability manager	**HUC 1,12 and 17**	With User/Access management
TC_3, processing, TC_13, reprocessing	Workflow, metadata, interoperability manager	**HUC 3,11 and 13**	Cross e-RI processing
TC_11, semantic linking **TC_16, reference model**	Metadata, interoperability manager	**HUC 2**	Cross e-RI linking, exploration and annotation

***Table 8.2.2.1**: ENVRIPLUS use-cases coverage by VRE4EIC*

9 Use cases evaluation *

Use case evaluation has been described in D2.5 (confidential deliverable). Many events took place to evaluate use cases. The main findings were derived from events 4, 5, 6 and 8 as described in D2.5. We summarize the main findings below.

In event 4 (January, 2017) we described evaluation of architecture using a SWOT analysis. TU Delft prepared a protocol for the evaluation events in consultation with ERCIM, INGV and UvA. Use cases were evaluated at project meetings of the EPOS project and the ENVRIPLUS project, since we wanted to ensure the participation of a group of highly experienced VRE developers. A weakness that was identified using the SWOT analysis was that the use-cases are generic and need to be "operationalised" to be tested. At the same time, there were opportunities for making the use cases more specific. Two opportunities were described in the workshop: HUC001 (one of the use cases) is applicable to the EPOS context and can thus be tested and one additional high level use-case can be added to VRE4EIC (note: this high-level use-case was never added, as we found a way to express it using two others high-level use-cases) by combining existing UC, so that the HUCs can be applicable to the ENVRIPLUS context. This is related to the threat that if only the applicable High level Use Cases (HUCs) are tested, e-VRE won't provide much value to EPOS and ENVRIPLUS; added-value can be brought to these communities by getting them interested to high-level use-cases they do not cover yet. The workshop participants stated that the use-cases already cover the full research lifecycle, and many low-level use cases have already been developed that may be more applicable in the context of EPOS and ENVRIPLUS.

In event 5 a two hour workshop was organized on December 19, 2016. This workshop was integrated in a project meeting for the EPOS IT team concerning work package 6 and 7 in Utrecht, The Netherlands. INGV led and implemented the evaluation event at the EPOS project meeting, which was also attended by TU Delft for assistance. During the meetings, Daniele Bailo (INGV) explained the use cases and the architecture from VRE4EIC and EPOS. He discussed the system architecture of the e-VRE together with the common operations needed/ provided. A brief SWOT analysis has been performed with representatives from EPOS. In this event, the participants stated that the method used to create the architecture and use cases is canonical, well tested and documented. Moreover, they stated that the research life-cycle use case does not cover reproducibility and reproducibility was added.

In event 6, a workshop was organized in January 2017. This workshop was integrated in a meeting for ENVRIplus RI representatives during the EUDAT user forum in Helsinki. UvA led and implemented the evaluation event with the ENVRIPLUS project members. During the meeting, Zhiming Zhao (UvA) explained use cases from VRE4EIC and ENVRIPLUS together with joint cases between ENVRIPLUS and EUDAT. He then discussed the system architecture of the e-VRE together with the common operations needed/provided by ENVRIPLUS and EUDAT. Based on the discussion, a SWOT analysis has been performed with representatives from ENVRIPLUS, ICOS, EISCAT and EUDAT. This event showed that the use cases nearly cover the entire lifecycle, (one experiment management HUC could be included, but was finally derived from the association of two other HUCs). This event also showed that the use case from the e-VRE can well complement the use cases from ENVRIPLUS (O1), which provides a good opportunity for RI developers and users to come to understand them.

In event 8, use case evaluation took place with TU Delft students. In March 2017, approximately 20 students of TU Delft participated in the evaluation of use cases. Students were first presented with the objectives and achievements of the VRE4EIC project. The already developed visionary use cases were presented to them and then the students were asked to create new visionary use cases. Each of these visionary use cases was discussed during the lecture. At the end of the lecture, the teacher summarised the results and proposed action points to mitigate weaknesses and threats of the use cases and to amplify their strengths and opportunities. After the lecture, we mapped the new visionary use cases to the existing VRE4EIC high level and low level use cases. This did not result in any new use cases, but it confirmed that the uses cases that were already developed in the project met all the requirements of the newly developed visionary use cases.

More details about the use case evaluation can be found in D2.5.

10 Validation of architecture design *

Use cases are the modelling technique for formalizing the functional requirements (Fun) of a system, the use cases analysis is the base upon which the system will be built. Functional requirements and use cases are in such way an approach to drive architectural decisions and can be used to validate the architecture.

It is important that the architecture supports the functional requirements of the current system.

This analysis has been conducted by retrieving the use-cases corresponding to the functions of the building blocks, with regards to the traceability matrix and the requirements table provided in the appendix of deliverables D2.3 and D3.1[5].

This method provides excellent traceability from the original requirements to the tested system.

The first step was to identify which functions are currently implemented in the building block, and the second step was to map these functions regarding the query requirements table, to be able to retrieve the list of use cases supported by the current eVRE prototype.

We have 59 use cases (UCs), 5 generalised functions (GFs) and 24 functions (Funs) identified and described for the project. All needed information about the functionalities and the traceability are available in D3.1. All the requirements gathered during the analysis phases, the use-cases and the high-level use-cases have been mapped to functional requirements, which comfort us in saying that the architecture design is aligned with the collected needs.

To go one step further, by using this traceability, we can know the percentage of original use-cases that are completed, which could be considered as an indicator of the completion of the prototype.

Here is the list of the generalised functions already implemented (functions marked with * are only partly implemented) in our eVRE prototype, with their corresponding elementary functions:

- **GF1 Query Service:** Fun1, Fun2, Fun3, Fun11*, Fun19
- **GF2 Dataset/Metadata Exploration:** Fun4, Fun5, Fun6, Fun11*, Fun19
- **GF4 Data Cataloguing:** Fun13, Fun14, Fun17, Fun18
- **GF5 Workflow Enactment:** Fun 15*
- **Node Service:** Fun20, Fun21, Fun22
- **App service:** Fun 23

After collecting these implemented function, we built a table with all the gathered requirements, the corresponding functions and the completion of these functions. The full table is available in annex B. This table also indicates the use-cases that the requirements are linked to. By checking the number of requirements for each use-case and the number of requirements that are completed, we can measure the level of completion of the use-cases.

In the first period of the project, a gap analysis has been conducted (see confidential deliverable D3.2). Priorities have been set on components regarding gap in functionalities analyzed in various VREs. Building blocks have been developed according to these priorities, which leads to the following results about completion of use cases. Please note that these results do not give any information about development workload, it only concerns the use-cases expressed by the users.

[5] https://www.vre4eic.eu/images/Public_deliverables/D3.1_Architecture_Design.pdf

Use cases concerning discovery, contextualisation and interoperability among assets of e-RIs using e-VRE (the major objective of the project) have been met to a level of 88%.

Use cases related to advanced data processing (scientific transformation, etc.), research and collaboration management (such as funding information) and equipment/sensor management were not prioritised in the development of the project and attained 10%.

11 Conclusion *

This deliverable shows the various steps taken to finalize the use-case analysis. It describes the way and methodology used to define use cases from requirements to visionary scenarios, and how this step is important for the system described by the different relations involved (see chapter 2). Visionary use-cases help us see realistic scientific scenario that need a VRE to be accomplished.

This deliverable also presented how the project aligns with EPOS and ENVRIPLUS high level us e cases, thus proving the alignments with the needs of different scientific communities.

By looking at the traceability matrix that links technical functions, requirements, use -cases, high-level use-cases and visionary use-cases, we also were able to validate the architecture and give a first indicator about the level of completion of the prototype, regarding the use -cases gathered.

12 References *

Ottinger, T., Langr, J. (2011). Agile in a Flash.

13 Annexes *

13.1 Annex A: Visionary use-cases description

VUC 1: Investigating historical interest based on current events

Discipline

Discipline: digital humanities.

Short case explanation

A digital humanities scholar would like to investigate whether the Great Depression in Europe from 1929-1940 is being studied as a reaction to the onset of the financial crisis of 2008. In order to do so, the researcher wants access to (historical) usage logs of different e-RIs containing data from this period and to articles and papers published on the subject of the Great Depression around 2008.

In order to do this, not only the right e-RIs need to be found, also relevant statistics on the transaction logs of 2008 and around of these e-RIs are needed. Also, papers written by other researchers and articles written by journalists about this period (both around 2008 and before and after) are of interest.

Possible relevant e-RIs:

- Delpher news archive (National Library of the Netherlands)
- Beeldengeluid (The Netherlands Institute for Sound and Vision)
- DANS (Data Archiving and Networked Services)
- NARCIS (National Academic Research and Collaborations Information System)
- CLARIAH (Common Lab Research Infrastructure for the Arts and Humanities)
- EUDAT (Research Data Services, Expertise & Technology Solutions) /Europeana Newspapers
- Google Scholar (papers/ log records both)
- another European REs

A junior researcher will first do some exploration, collecting some basic statistics based on differential privacy [1,2,3] (differential privacy is a framework used to protect against a privacy breach by e.g. combining datasets).

After the initial collection of information, a senior digital humanities researcher will perform a more in-depth analysis of the interest in the Dutch Depression, or the 'crisis jaren'. For this part, the focus is on the Netherlands only and the researcher will need more detailed (and privacy sensitive) information regarding the usage logs.

Use-case description

As a <junior digital humanities researcher> I want to <analyse usage logs from different REs> in order to <discover whether there is a renewed interest in the Great Depression in Europe between 1929-1940 since the financial crisis has hit the EU>.

As a <senior digital humanities researcher> I want to <analyse usage logs of the Dutch REs> in order to <discover where interest is focused on with respect to the Netherlands and where this interest is originating from>.

Actors involved in the use case

System users

- junior digital humanities researcher (without NDA, authorized by AAAI/CERIF)
- senior digital humanities researcher (with NDA, authorized by AAAI/ CERIF, project member of a relevant EU project)

Priority

High

Pre-conditions

The users are signed on via EPOS/CERIF.

Flow of events – user view

1. Junior researcher ticks the box to give permission to use CERIF data for the single sign on and all further authentication
2. Junior researcher requests list of relevant e-RIs in the EU providing publications of the 1930's and their corresponding transaction logs
3. Junior researcher saves this list
4. Junior researcher requests basic usage statistics computed on the relevant e-RIs:
    a. (Aggregated) statistics on usage logs of access to documents where metadata matches:
        i. Time period
        ii. Subject: financial reporting
    b. Metadata statistics on relevant accessed documents
5. Junior researcher receives and stores the outcome of the relevant Res
6. Junior researcher requests list of relevant e-RIs in the EU providing information on publications (newspaper articles and research papers) in 2008 and around
7. Junior researcher saves this list
8. Junior researcher requests basic statistics from these e-RIs
    a. about articles and papers published in 2008
    b. with the metadata subject matching the Great Depression in the EU
9. Junior researcher receives and stores the results and signs out
10. Senior researcher ticks box to give permission to use CERIF data for the single sign on as a project member of an EU project
11. Senior researcher refines the list of the relevant e-RIs created by the junior researcher to only those e-RIS in the Netherlands providing publications of the 1930's and the corresponding transaction logs
12. Senior researcher requests more direct access to the relevant usage logs based on a previously signed NDA and membership of the EU project
13. Senior researcher performs a more in-depth analysis of the transaction
    a. e.g. asking where the interest in these documents originated from (location of user)
14. Senior researcher receives and stores the results

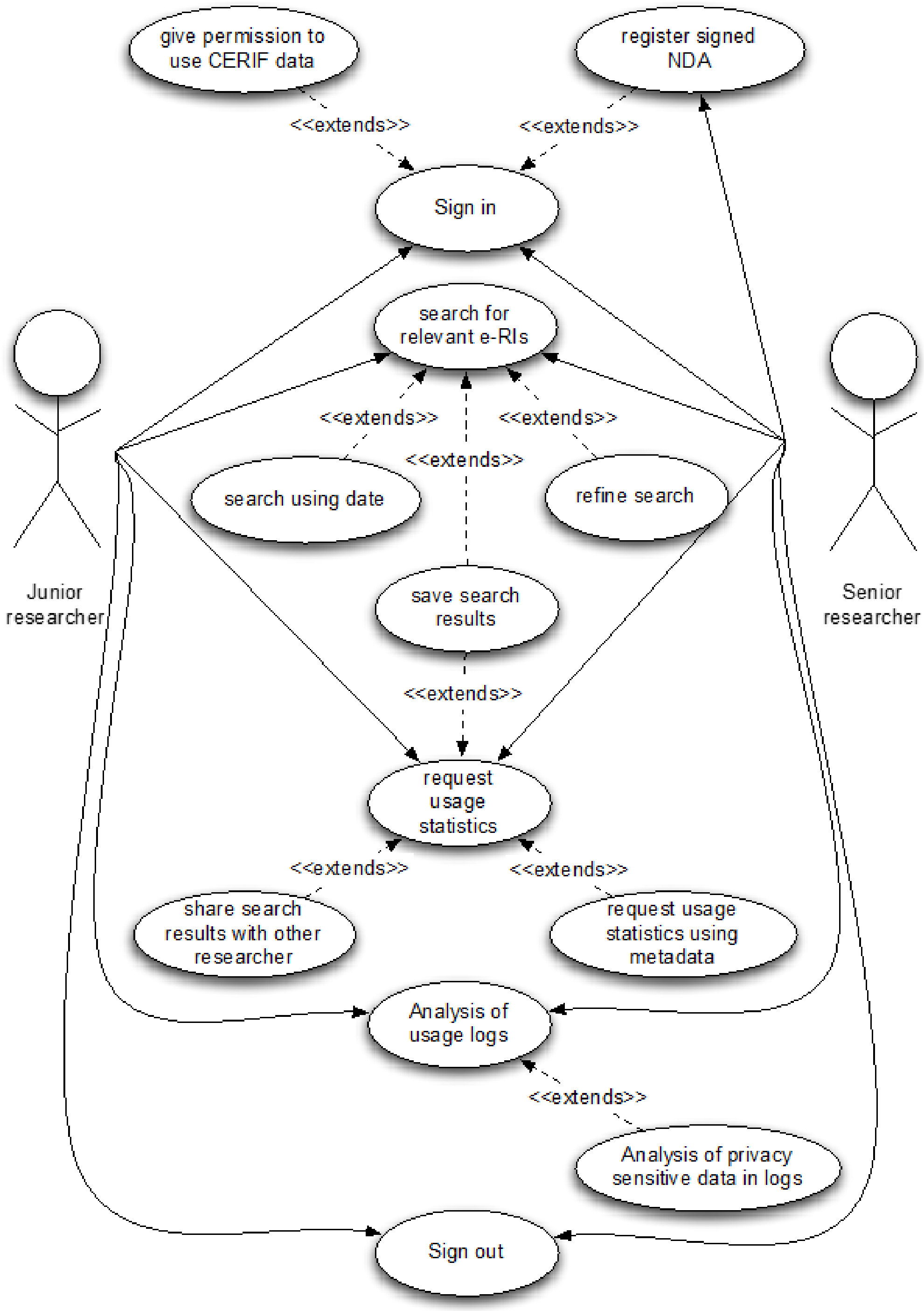
System workflow - system view

1. System uses a SAML based SSO (e.g., Unity-idm) to verify the identity of the junior researcher and retrieves the CERIF roles of this researcher
2. System searches eRI metadata to find e-RIs that contain access logs to digital libraries such as newspaper archives
3. System broadcasts the researchers query to the selected e-RIs, providing each e-RI with the user's identity and CERIF access clearance encoded as a signed SAML message
    a. Each e-RI validates the SAML message and verifies that the researcher has the right to access the data at the requested level of aggregation. It either replies with the requested statistics or a permission error.

4. The above procedure is repeated for 2008
5. System correlates the data retrieved from the two searches, e.g., using R
6. Interaction on behalf of the senior researcher is the same as for the junior researcher, except that the authentication profile carries and attribute that states that the researcher has signed an NDA and the e-RIs use this information to grant access to the log data at a lower level of aggregation

Post-conditions

- Researchers are satisfied with the results.
- Results are stored on request for the researcher.
- Privacy has been preserved.



References

1. Dwork, C., 2006. Differential privacy. In *Automata, languages and programming* (pp. 1-12). Springer Berlin Heidelberg.

2. Dwork, C., Rothblum, G.N. and Vadhan, S., 2010, October. Boosting and differential privacy. In *Foundations of Computer Science (FOCS), 2010 51st Annual IEEE Symposium on* (pp. 51-60). IEEE.

3. Arik Friedman and Assaf Schuster. 2010. Data mining with differential privacy. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining* (KDD'10). ACM, New York, NY, USA, 493-502. DOI=http://dx.doi.org/10.1145/1835804.1835868

VUC 2: investigating capacity of urban infrastructure in case of evacuation

Disciplines

Geological, mobility, sociological and housing.

Short case explanation

This scenario describes a (governmental) researcher performing research to the sufficiency of the capacity of infrastructure near cities in case of an evacuation. The evacuation may be due to a natural disaster (volcano explosions, floods). Understanding the weaknesses and undercapacity in the urban infrastructure may prevent casualties in case of an emergency evacuation.

The researcher would like to obtain data regarding the frequency and impact of (for example) volcano eruptions over the last X years in country Y. This knowledge is used to calculate the risks involved in future eruptions. The ability to estimate the amount of possible area that will be damaged by such disasters, can improve evacuation plans for citizens and result is less casualties in case of a disaster.

The volcanoes can be plotted on a map using the OpenStreetMaps API. A heat map visualization is able to show the locations of the volcanoes and the impact of the volcano explosions over the last X years. The volcanoes are dots on a map, the impact is visualized using the colors of the heat map. More intense means more impact.

Using this OpenStreetMaps visualization of volcanoes, the researcher is able to locate nearby cities C that fall within the range of the impacted area. The researcher the searches for datasets of a particular nearby city Ci, containing infrastructure information, topology and population density information of Ci. This data should provide insight in the population density of the impacted area and the capacity of the nearby transport infrastructure to evacuate.

Plotting the infrastructure and population density information on an OpenStreetMaps visualization, the researcher is able to perform an estimation regarding the capacity of the available infrastructure, based on the intensity of the eruptions. The visualization of the OpenStreetMaps can again be a heat map that visualizes the denser parts of the area. The researcher is then able to perform a calculation (offline) which results in an estimation of how much people would need to be evacuated and the amount of time necessary to accomplish that evacuation.

Based on the information regarding the amount of people that need to be evacuated, the researcher searches for a dataset regarding housing availability in cities near city Ci. Available housing accommodation is interesting to know in case of an emergency. Casualties may be able to make use of the available housing in other nearby cities. The researcher is able to request a map from the system showing the available housing spots per nearby city.

Use-case description

As a <governmental research> I want to be able to <view and visualize on an Open Street Maps map, a dataset regarding the frequency and impact of volcano explosions and plot in the same map a dataset regarding the urban/transport infrastructure in a certain area and see citizen counts related to nearby

cities on the same map> to that I can <gain an understanding of capacity and level of sufficiency of the available urban/transport infrastructure and housing in nearby cities in case of an emergency>.

Actors involved in the use case

Governmental researcher

Priority

Medium

Pre-conditions

User must have logged in

Flow of events – user view

Basic sequence and needed steps

1. Obtain data regarding the frequency and impact of (for example) volcano explosions over the last X years in country Y and visualize this in a heat map.
2. Locate impacted cities C by a volcano eruption.
3. Obtain population density data and urban/transport infrastructure data of city Ci.
4. Plot the infrastructure and population density information on an OpenStreetMaps visualization.
5. Based on capacity facts of the infrastructure, an estimation of the level of sufficiency of the capacity of the available urban infrastructure can be made.
6. Based on the impact of the volcano eruption, the researcher is able to estimate the amount of casualties after an eruption and using a data set containing housing accommodation in nearby cities, estimate whether there are enough housing accommodations in other cities.
7. The researcher is able to download the results or share the results with other users of the system.

System workflow - system view

1. The UI (User Interface) receives the input: “volcanoes” and “country”
2. Backend system connects to the database and API’s and searches for datasets that have those input parameters in its title.
3. The backend retrieves the datasets matching the input criteria and represents this list of possibilities to the user.
4. The UI receives the selected dataset entry and requests the back-end to fetch the entire dataset from the database and loads this onto the client.
5. The UI receives the action to visualize the columns “volcano name”, “coordinates”, “level of impact” and “coordinates of the lava flood area”.
6. The system recognizes coordinate format, so it retrieves the corresponding area from the OpenStreetMaps API.
7. OpenStreetMaps canvas is pained using the data points from the input columns, which results in a heat map.
8. The painted heat map, without sacrificing OpenStreetMaps features and details, is served to the user. Types of urban infrastructure should thus be still visible in the plot.
9. The user is able to, in a “tabbed-window”, create a new workspace in which the user searches for a dataset regarding the population density information. The system again recognizes the dataset that contains geolocation data points and provides the option to the user to visualize this in a map.
10. The user is then asked by the system through means of a UI dialog whether the existing plot should be used to visualize the population destiny information.

11. The user selects “Yes” to the dialog presented in 10, and is served with a plot in which both the volcano impact and population density data sets are plotted, by means of a heat map visualization.
12. The user searches for a dataset regarding housing in nearby cities of city C-I. The system asks the user if he/she wants to use an existing plot.
13. The user answers “Yes” to the dialog in 12 and selects the columns of the dataset that the user wants to plot in the existing map.
14. The system receives the columns “Address”, “availability”, “amount” and plots these in the existing map.
15. The system shows in the same heat map plot as in 7 the numbers of available housing accommodations.
16. Based on the observations, the user downloads the results for offline use. In addition, the user is able to share the results with other users of the system.

VUC 3: Evolution of electric charging

Disciplines

Urban infrastructure, mobility, power grid topology.

Short case explanation

A (governmental) researcher wishes to explore the availability of charging points for Electrical Vehides (EV), so that an estimation can be made regarding the sufficiency of the availability of charging points with respect to the amount of EV’s in a particular area. The governmental researcher searches for a dataset containing the locations and availability for charging points of EU country A. The researcher is now able to visualize the locations of the charging points using OpenStreetMaps.

Subsequently, the researcher wants to plot data regarding the number of households in a city in procession of an Electrical Vehicle on the same map. Based on that knowledge, the researcher is able to estimate the sufficiency of the available charging points.

The researcher is able to obtain an overview of the relations between EV's and available public charging points in country A and assess whether this is sufficient. In places where the amount of charging points is insufficient, the researcher wants to gain insight in the possibilities to increase this amount. In order to do so, the researcher first searches for possible physical limitations. Therefore, the researcher searches for the power grid topology of the city and its capacities. Based on these data, the researcher is able to advice the government where possible charging spots can be added and where the infrastructure allows this expansion.

Use case description

As a <(governmental) researcher> I want to be able to <view and compare the availability of charging points in an EU city/country A and the number of electric vehicle (EV) in A> so that I can <gain an understanding of the level of sufficiency of the number of charging points in A and the ability to increase the amount of charging points>.

Actors involved in the user case:

System user

- (governmental) researcher

Priority

Medium

Pre-conditions:

The use must have logged in.

Flow of events – user view

1. <researcher> chooses country/city of interest.
2. <researcher> plots the geolocation points of charging points in a map.
3. <researcher> plots the geolocation of EV registered vehicles in the same map.
4. <researcher> performs an analysis based on the visualization.
5. <researcher> searches for power grid topology and plots this in the map.

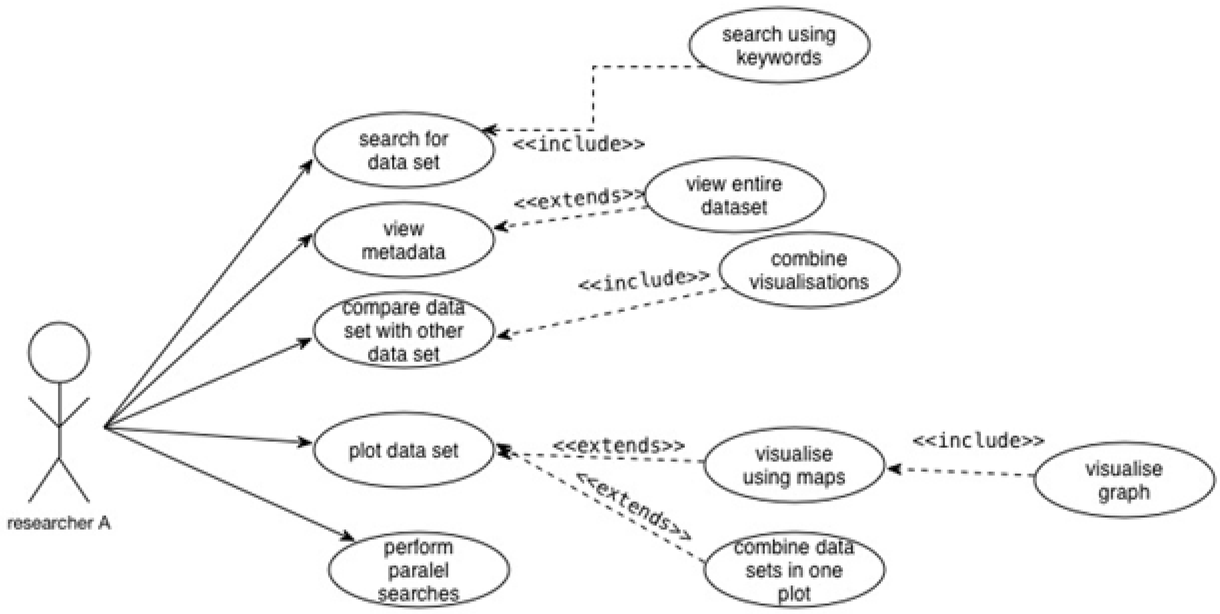
System workflow – system view

1. The UI receives the input "EV" and "<country and city name A>"
2. The system connects to the database and connected systems to request datasets containing EV and <country and city name A> in the title or metadata of the dataset.
3. The system presents, via the UI, a list of datasets it retrieved from the database to the user.
4. The user selects the dataset it is interested in and the system shows possible interaction with the dataset. Possible interactions are: "view dataset", "plot dataset in chart" or "plot dataset on map".
5. The user selects the "plot on OpenStreetMaps"-interaction, the system recognizes the format of the geolocation coordinates and presents these columns to the user.
6. The user selects the columns "location" and "amount" and the filled in map is represented to the user.
7. The user is able to, in a "tabbed-window", create a new workspace in which the user searches for a dataset regarding the number of EV's in <country and city name A>. The system again recognizes the dataset that contains geolocation data points and provides the option to the user to visualize the EV dataset in a map.
8. The system shows the possible columns: "zip code" and "amount".

9. The user is then asked whether the existing plot should be used to visualize the EV information.
10. The user selects "Yes" and is served with a plot in which both the charging points and number of EV's data sets are plotted. The user is able to see green coloured dots that represent charging points and orange coloured dots that represent EV's. In addition, an amount is provided per cluster of dots.
11. The system's search field receives input "power grid" and "city A" and searches in its connected system for a dataset matching those keywords.
12. The system provides a list of possible matches to the user.
13. Based on the selection of the user, the system loads the underlying dataset from its connected system and recognises the geometrical data points that represent a graph.
14. The asks the user it wants to plot the power grid graph in the same map plotted in 5.
15. The user selects "Yes" on this dialog.
16. The system draws the graph on the map and represented this map, including all the drawn objects to the user.

Post-conditions

- The user is served with a single OpenStreetMaps map in which both data sets are plotted.
- The Open Street Maps visualization is in a separate window, to ensure flexibility an integration with other personal workflows of the user.



VUC 4: Relation between age of first pregnancy and level of education followed

Disciplines

Healthcare, sociology, education

Short case explanation

A researcher would like to investigate in the relation between the age of first pregnancy and the level of education under women in a certain country A. Understanding this relation may contribute to improvements in the education system in country A.

The researcher first searches for a dataset that contains information of the first pregnancies of women. Such a dataset may be structured according to age ranges with a corresponding percentage of pregnancies. Then, together with that dataset, the researcher searches for a dataset regarding the level of education of women in a certain age range. Based on that information, the researcher is able to find high level relations between the amount of pregnancies at a certain age and their (presumed) corresponding level of education. The researcher is able to select rows and columns in both datasets and plot the data in a bar chart so that different age categories are easy to compare.

Use case description

As <a researcher> I want to be able to <view and compare the age of first pregnancy under women and the level of education in a EU country> to <gain insights in the relation between age of first pregnancy and level of education>.

Actors involved in the user case

System user

- Researcher

Priority

Medium

Pre-conditions

The use must login.

Flow of events – user view

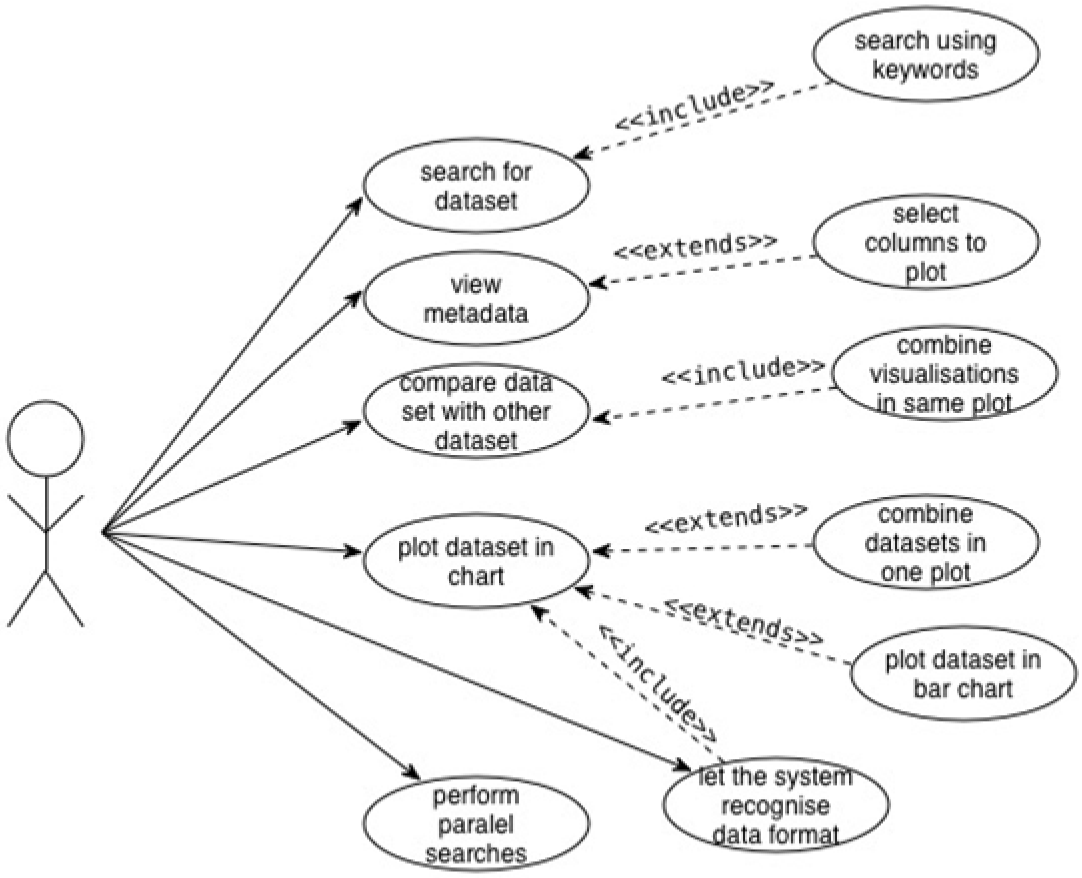
1. <researcher> chooses country of interest.
2. <researcher> plots the data points of age of first pregnancy in a bar chart plot.
3. <researcher> plots the data points of level of education in the same a bar chart plot
4. <researcher> performs an analysis based on the visualization.

System workflow – system view:

1. The UI receives the input "first pregnancy age women" and "<country name A>"
2. The system connects to the database and connected systems to request datasets containing "first pregnancy age women" and "<country name A>" in the title or metadata of the dataset.
3. The system presents, via the UI, a list of datasets it retrieved from the database to the user.
4. The user selects the dataset it is interested in and the system shows possible interaction with the dataset.
5. The user is able to view the metadata and columns in the dataset.
6. The user selects the "plot in chart diagram"-interaction, the system recognizes the format of the data and asks the user in which type of chart the selected columns have to be plotted.
7. The user selects a bar chart graph and the system retrieves the required data from the underlying database system.
8. The user is able to, in a "tabbed-window", create a new workspace in which the user searches for a dataset regarding the level of education under women in <country and city name A>. The system recognizes similar data points as plotted in the bar chart and provides the option to the user to plot the education data set in the same chart.
9. The user selects "Yes" and is served with a bar chart in which both the age (percentage per age range) of first pregnancy under women and the level of education in country A data sets are plotted.

Post-conditions:

- The user is served with one or multiple windows in which the plots are visualized.
- The visualizations are in a separate window, to ensure flexibility an integration with other personal workflows of the user.
- The user is able to view the underlying raw dataset.



VUC 5: Social consequences of increased debts of EU countries

Disciplines

Finance, economic and criminology

Short case explanation

An economist (researcher) would like to investigate in the relation between the governmental debts with respect to their GDP and unemployment and criminality rate of EU countries over the last X years. Insights in these relations may contribute to understanding of economical mechanisms and the increase in criminality.

The researcher searches for datasets containing governmental debts with respect to the GDP of EU countries. The researcher can select from datasets the key figures that it wishes to plot using a graph diagram. In this graph visualization of the key figures, the researcher has the ability to perform basic regression analysis of the key figures. Based on the results of the analysis, the researcher wants to understand two things:

- How does this affect the unemployment rate in that country?
- How does this affect the criminality rate in that country?

To facilitate the researcher in finding relations between the data points, the system provides some basic analysis tools. The researcher makes use of linear regression to extend the relationship between the three components in the bar chart.

Insight in the relation between debts of a country, unemployment rate and criminality rates in that country, may lead to a better understanding of the social effects of unemployment in a poorer country. Thereafter, the researcher is able to export the used data in CSV/MS Excel format.

Use case description

As <a researcher> I want to be able to <view and compare the GDP of country X and Y and their unemployment rate respectively> to <gain insights in the relation between the debts with respect to their GDP and unemployment rate and the criminality rates>.

Actors involved in the user case

System user

- Researcher/economist

Priority

Medium

Pre-conditions

The use must login

Flow of events – user view

1. <researcher> chooses country X of interest.
2. <researcher> plots the data points of debts of X in a bar chart plot.
3. <researcher> plots the data points of unemployment rate of X in a bar chart plot.
4. <researcher> plots the data points of criminality rate of X in a bar chart plot.
5. <researcher> performs an analysis based on the visualization.
6. <researcher> downloads the visualization.

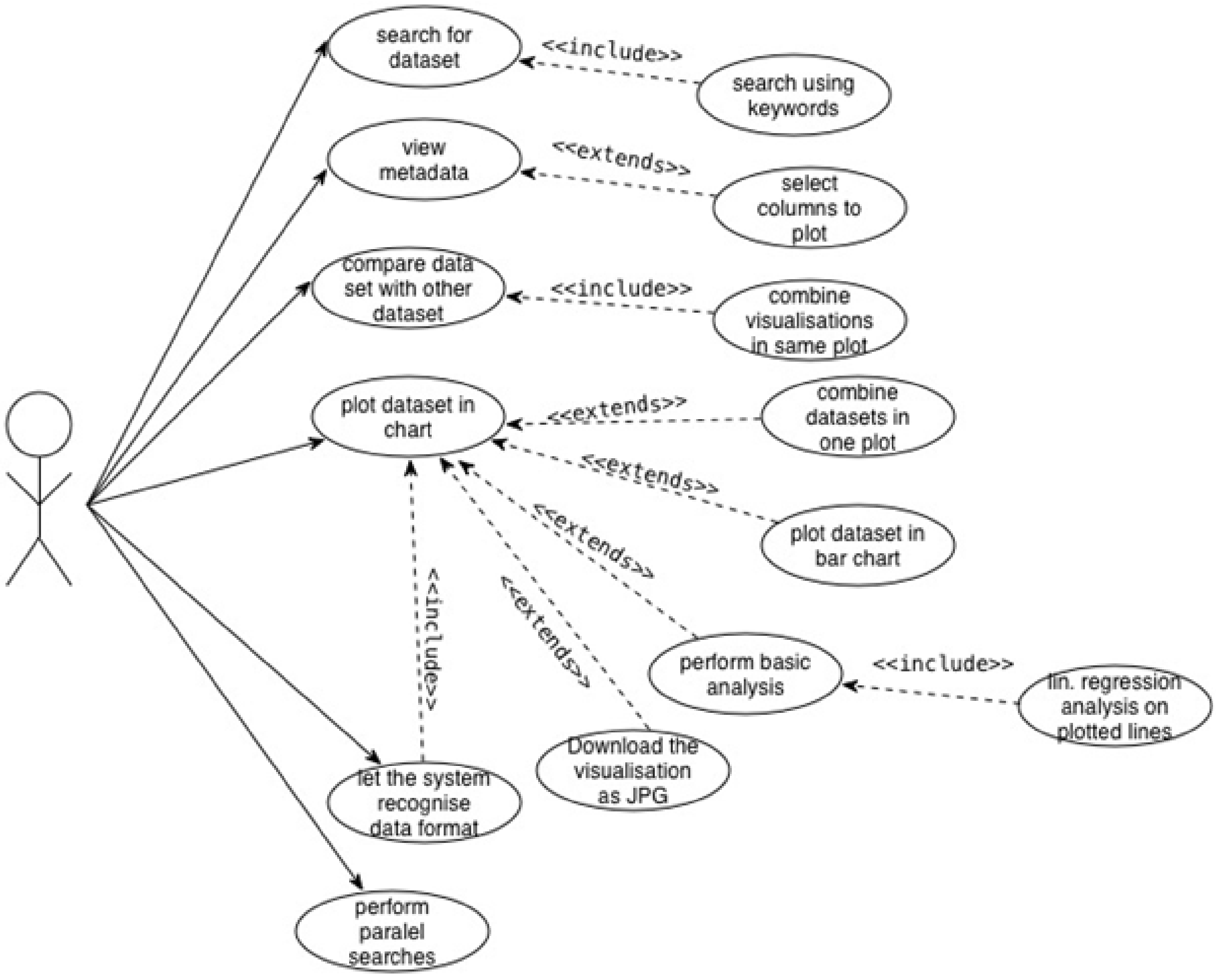
System workflow – system view

1. The UI receives the input "debts" and "<country name X>"
2. The system connects to the database and connected systems to request datasets containing "debts" and "<country name X>" in the title or metadata of the dataset.
3. The system presents, via the UI, a list of datasets it retrieved from the database to the user.
4. The user selects the dataset it is interested in and the system shows possible interaction with the dataset.
5. The user selects the "plot in chart diagram"-interaction, selects the columns the user is interested in.
6. The system recognizes the format of the data in the selected columns of the dataset and plots the data points in a bar chart.
7. The user is able to, in a "tabbed-window", create a new workspace.
8. In this new workspace, the UI receives "unemployment rate" and "<country X>".
9. The system searches the database and connected systems and presents the user with a list of matching datasets.
10. The system gets input regarding the selected dataset and shows the available interactions with the dataset.
11. The system receives "bar chart" input and presents the user with a bar chart in which both the debts and unemployment rate in country X data sets are plotted.
12. The user is able to, in a "tabbed-window", create a new workspace.
13. In this new workspace, the UI receives "criminality rate" and "<country X>".
14. The user selects the dataset and the interested columns "total amount of offences in year".
15. The user asks the system to plot this data in a bar chart, in the same plot as constructed by the system in 6. The user is then able to derive the relation between unemployment rate and criminality, based on the bar chart.

16. Besides that, the user uses one of the basic analysis tools in the menu of the bar chart: regression analysis.
17. The system performs a linear regression analysis on the 3 different data sets in the bar chart and plots these in the bar chart visualization.
18. The user clicks on “Download as JPG” and the system exports the visualization to JPG and uploads it to the user’s client.

Post-conditions

- The user is logged in
- The user is served with one or multiple windows in which the plots are visualized.
- The visualizations are in a separate window, to ensure flexibility an integration with other personal workflows of the user.
- The user is able to view the underlying raw dataset.



VUC 6: Shipwreck archaeology

Disciplines

Environment, engineering and social science(history)

Use case description

As a <earth science researcher> I want to <collect remote sensing data from satellites, such the images of earth surface> to <store in the e-VRE>.

As a <marine environment researcher> I want to <download the historical marine tide data> to <test my tide flow model>.

As a <marine engineering researcher> I want to <use the images captured by satellite and tide and current simulation data> to <design a modelling method to detect the location of the shipwrecks close to coastal line>.

As an < archaeologist> I want to <use the location data of shipwrecks > to <conduct underwater archaeological excavation at more accurate location>.

Actors involved in the user case:

System User

- <Earth science researcher>
- <Marine environment researcher>
- <Marine engineering researcher>
- <Archaeologist>

Priority

High

Pre-conditions

- Users must login
- The e-VRE allows for data transmission and storage.

Flow of events – user view:

1. <Earth science researcher> gets the images of the coastal line from the satellite.
2. <Earth science researcher> uploads the image of the coastal line to the e-VRE.
3. <Marine environment researcher> uploads the flow simulation of tides and currents to the e-VRE.
4. <Marine engineering researcher> downloads the image of the sea and tidal simulation data.
5. <Marine engineering researcher> designs modelling method to detect the location of shipwrecks by using abovementioned data.
6. <Marine engineering researcher> uploads the location data of shipwrecks to the e-VRE.
7. <Archaeologist> sends a request to the Marine engineering researcher to access to the shipwreck location data.
8. <Marine engineering researcher> approves the data request from the archaeologist.
9. <Marine engineering researcher> rejects the data request from the archaeologist
10. <Archaeologist> downloads the location data and design scenario for underwater excavation.

System workflow – system view:

1. In order to execute step 1, e-RI underline the e-VRE needs to communicate with satellite and store the image
2. In order to execute step 2,
3. e-VRE needs to be able to recognize the metadata of the images and store the image.
4. In order to execute step 3, e-VRE needs to be able to recognize the metadata of the tidal simulation file and store the simulation file.
5. In order to execute step 4, e-VRE needs to communicate with users' devices and allow user to search for the specific data.
6. In order to execute step 5, e-VRE provides modelling tools.
7. In order to execute step 6, e-VRE needs to recognize the metadata of the shipwreck location file and store the location file.
8. In order to execute step 7, e-VRE provides Marine engineering researcher with data management functionalities.

9. In order to execute step 8, e-VRE needs to trigger and execute data request workflow.
10. In order to execute step 9, e-VRE needs to communicate with user’s computer.

Post-conditions:

- Satellite images are stored in the e-VRE,
- Tidal simulations are stored in the e-VRE.
- There are usage logs on who upload, download and change the versions of the data.
- There are usage logs on how the data publisher defines the data access rights.

VUC 7: Prediction of transport delay

Disciplines

Environmental sciences, transportation

Short case explanation

A (transportation) researcher wishes to build a model for prediction of transport delay in a country (or region) in the certain period of the year. The model should represent the correlation between environmental (weather) and transportation data. Dataset with georeferenced and time-tagged weather data and transportation data are necessary for building this model.

The researcher searches for a dataset containing the locations; dates and times; and weather data for the certain country or region. Subsequently, the researcher searches for a dataset containing the transportation and traffic flow/speed/congestion data for the certain country or region - GPS Vehicle Travelling Recorder with Digital Tachograph, roadside detectors which can be used for monitoring traffic flow/speed/congestion, websites showing speed of traffic flow on sections of roads developed by some European countries highways agencies can be used as source for these data.

Based on these datasets, the researcher is able to find relations between weather and transportation data and improve existing model of correlation between weather data and transportation delay. The researcher is able to select rows and columns in both datasets and plot the data in a bar chart so that different weather conditions are easy to compare. This improved model can be used for prediction of transportation delay using weather forecast data and real-time transportation data. Furthermore, the prediction can be used for better organization of transportation companies including the organization of connected public transports. Also, if transportation delays in some company or in some region are not in accordance with predicted delays, the company should analyse whether the company’s vehicles should be better equipped for the certain weather conditions, or the government of the region should analyse whether the region’s transportation infrastructure should be improved.

Use case description

As a <(transportation) researcher> I want to be able to <view and compare travel time and weather conditions for some region> so that I can <gain insights in the relation between travel time and weather condition> in order to create a model which could better predict transport delay and better organize connected transports.

Actors involved in the user case

System user

- Transportation researcher

Priority

Medium

Pre-conditions

The user has to be logged in.

Flow of events – user view

1. <researcher> chooses country or region of interest.
2. <researcher> request a weather dataset.
3. <researcher> plots the data points of weather data in a bar chart plot.
4. <researcher> request a transportation dataset.
5. <researcher> plots the data points of transportation data in the same bar chart plot.
6. <researcher> request a traffic flow/speed/congestion dataset.
7. <researcher> plots the data points of traffic data in the same bar chart plot.
8. <researcher> performs an analysis based on the visualization.
9. <researcher> creates a model which represent the correlation between weather data, traffic data and transport time.
10. <researcher> uploads a weather forecast dataset.
11. <researcher> plots the data points of weather forecast data in a new bar chart plot.
12. <researcher> uploads a real-time transport data - position (GPS) and state (like if they are stationary or in an accident).
13. <researcher> plots the data points of real-time transport data in a new bar chart plot.
14. <researcher> uploads a real-time traffic data - roadside detectors which can be used for monitoring traffic flow/speed/congestion, websites showing the speed of traffic flow on sections of roads developed by some European countries highways agencies.
15. <researcher> plots the data points of real-time traffic data in a new bar chart plot.
16. <researcher> plots the data points of calculated transportation data in the same bar chart plot. Predicted values of transport times are calculated using a model created in point 7.

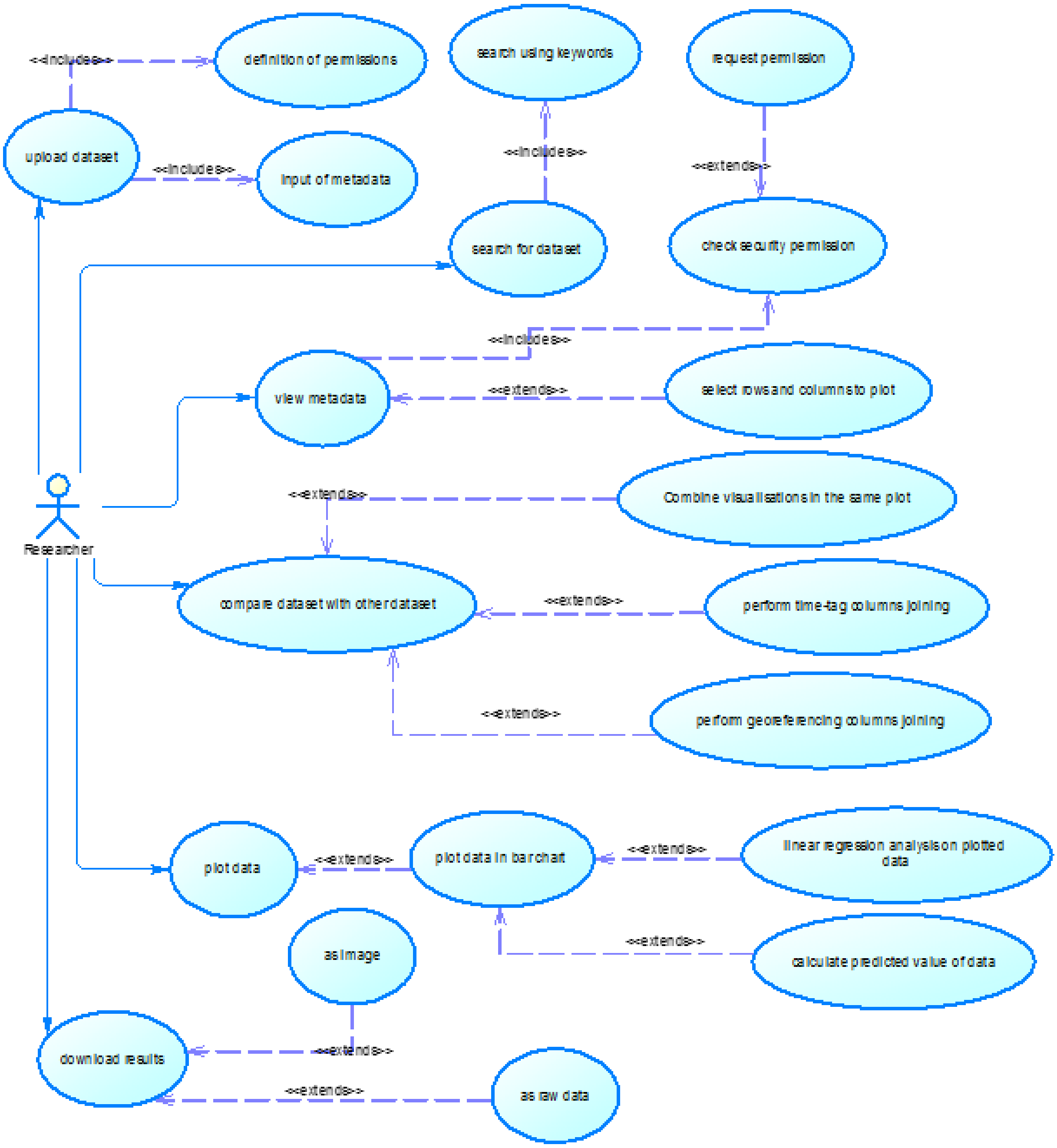
System workflow – system view

1. The UI receives the input "weather data" and "<country name X>"
2. The system connects to the database and connected systems to request datasets containing "weather data" and "<country name X>" in the title or metadata of the dataset.
3. The system searches the database and connected systems and presents the user, via the UI, a list of matching datasets.
4. The user selects the dataset he/she is interested in, the system checks whether access to the dataset is allowed to the user which is signed in, if does, the system shows possible interaction with the dataset, else the system send a dataset permission request to the owner of the dataset.
5. The user selects the "plot in chart diagram"-interaction, selects the columns the user is interested in, selects the columns which will be used as time-tag, selects the columns which will be used for georeferencing.
6. The system recognizes the format of the data in the selected columns of the dataset and plots the data points in a bar chart.
7. The user is able to, in a "tabbed-window", create a new workspace.
8. In this new workspace, the UI receives the input "transportation data" and "<country name X>".
9. The system searches the database and connected systems and presents the user with a list of matching datasets.
10. The user selects the dataset he/she is interested in, the system checks whether access to the dataset is allowed to the user which is signed in, if does, the system shows possible interaction with the dataset, else the system send a dataset permission request to the owner of the dataset.
11. The user selects the "plot in chart diagram"-interaction, selects the columns the user is interested in (including starting point, end point, transportation time), selects the columns which will be used as time-tag, selects the columns which will be used for georeferencing

12. The user asks the system to plot these data in a bar chart, in the same plot as constructed by the system in 6, and to group data using time-tag and georeferencing columns in step 5 and 11.
13. The user is able to, in a “tabbed-window”, create a new workspace.
14. In this new workspace, the UI receives the input “traffic data” and “<country name X>”.
15. The system searches the database and connected systems and presents the user with a list of matching datasets.
16. The user selects the dataset he/she is interested in, the system checks whether access to the dataset is allowed to the user which is signed in, if does, the system shows possible interaction with the dataset, else the system send a dataset permission request to the owner of the dataset.
17. The user selects the “plot in chart diagram”-interaction, selects the columns the user is interested in (flow, speed, congestion), selects the columns which will be used as time-tag, selects the columns which will be used for georeferencing
18. The user asks the system to plot these data in a bar chart, in the same plot as constructed by the system in 12, and to group data using time-tag and georeferencing columns in step 5, 11 and 17. The user is then able to derive the relation between weather data, transportation time and traffic data, based on the bar chart.
19. Besides that, the user uses one of the basic analysis tools in the menu of the bar chart: regression analysis.
20. The system performs a linear regression analysis on the 3 different data sets in the bar chart and plots these in the bar chart visualization.
21. The user is able to, in a “tabbed-window”, create a new workspace.
22. In this new workspace, the user uploads “weather forecast data” obtained from some other source for the certain region or country.
23. The user selects the dataset permission and metadata.
24. The user selects the “plot in chart diagram”-interaction, selects the rows and columns the user is interested in (same column as in 5), selects the columns which will be used as time-tag, selects the columns which will be used for georeferencing.
25. In the new workspace, the user uploads “real-time transportation data” obtained from some other source for the certain vehicles.
26. The user selects the dataset permission and metadata.
27. The user selects the “plot in chart diagram”-interaction, selects the rows and columns the user is interested in, selects the columns which will be used as time-tag, selects the columns which will be used for georeferencing.
28. In this new workspace, the user uploads “real time traffic data” obtained from some other source for the certain region or country.
29. The user selects the dataset permission and metadata.
30. The user selects the “plot in chart diagram”-interaction, selects the rows and columns the user is interested in, selects the columns which will be used as time-tag, selects the columns which will be used for georeferencing.
31. The user asks the system to plot all these data in the same bar chart, to group data using time-tag and georeferencing columns in step 24, 27 and 30, and to calculate transport time using the result of linear regression analysis in 20 and to plot this calculated data also.
32. The user clicks on “Download as JPG” and the system exports the visualization to JPG and the user can download that.

Post-conditions

- The user is served with one or multiple windows in which the plots are visualized.
- The visualizations are in a separate window, to ensure flexible integration with other personal workflows of the user.
- The user is able to view the underlying raw dataset.



VUC 8: Choosing travel destination for tourists with allergic diseases

Disciplines

Healthcare, environmental sciences

Short case explanation

A (healthcare) researcher wishes to build a model for developing a warning/recommendation system for tourists with allergic diseases for choosing travel destination in a country (or region) in the certain period of the year. The model should represent the correlation between environmental and healthcare data taking into account time and geospatial dimensions. Dataset with georeferenced and time-tagged environmental data and healthcare (allergic) data are necessary for building this model.

The researcher searches for a dataset containing the locations; dates and times; and environmental data for the certain country (or region). Subsequently, the researcher searches for a dataset containing the time-tagged (with time dimension) healthcare data (allergic data) georeferenced using medical institution address for the certain country (or region).

The researcher is able to select rows and columns in both datasets and plot the data in a bar chart so that different environmental data and allergic data for some regions and period of years can be easily compared. Based on these datasets, the researcher is able to find relations between the environmental and allergic data and improve existing model of correlation between environment and allergic diseases. This improved model can be used for prediction of allergic problems for a certain period of the year for regions where we have only environmental data (not healthcare – allergic data). That prediction can be visualized using OpenStreetMaps API, predicted allergic diseases can be plotted on a map using different colors for different allergic diseases and periods of the year, and size of the circle can represent the probability of allergic reaction. That map can be used by a tourist with allergic diseases for choosing travel destination for the certain period of the year.

Use case description

As a <(healthcare) researcher> I want to be able to <view and compare the environmental data and allergic dataset of a region> to <gain insights in the relation between the time of the year, weather and the allergy rates for the region> in order to create a warning/recommendation system for tourists with allergic diseases before choosing travel destination in the certain period of the year.

Actors involved in the user case

System user

- Healthcare researcher

Priority

Medium

Pre-conditions

The user has to be logged in.

Flow of events – user view

1. <researcher> chooses country or region of interest.
2. <researcher> request an environmental dataset.
3. <researcher> plots the data points of environmental data in a bar chart plot.
4. <researcher> request an allergic diseases dataset.
5. <researcher> plots the data points of allergic diseases data in the same bar chart plot.
6. <researcher> performs an analysis based on the visualization.
7. <researcher> creates a model which represent the correlation between environmental data and allergic diseases.
8. <researcher> plots the data points of predicted value in the same bar chart plot for missing allergic diseases data. Predicted values of numbers of allergic diseases are calculated using a model created in point 7.
9. <researcher> plots the data in the map.

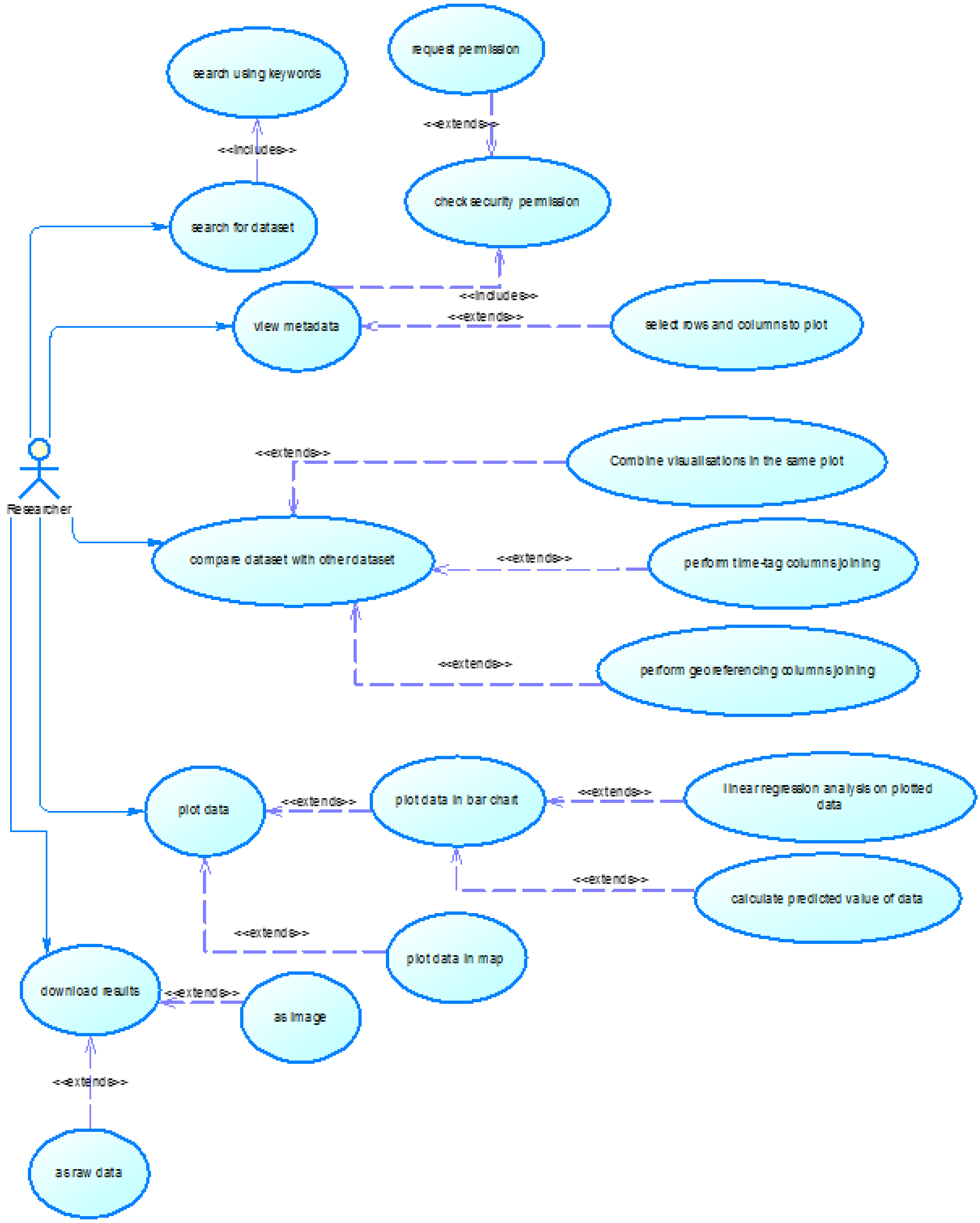
System workflow – system view

1. The UI receives the input "environmental data" and "<country name X>"
2. The system connects to the database and connected systems to request datasets containing "environmental data" and "<country name X>" in the title or metadata of the dataset.
3. The system presents, via the UI, a list of datasets it retrieved from the database to the user.

4. The user selects the dataset it is interested in, the system checks whether access to the dataset is allowed to signed in user, if does, the system shows possible interaction with the dataset, else the system send a permission request to the owner of the dataset.
5. The user selects the “plot in chart diagram”-interaction, selects the columns the user is interested in, selects the columns which will be used as time-tag (week, month or year season), selects the columns which will be used for georeferencing.
6. The system recognizes the format of the data in the selected columns of the dataset and plots the data points in a bar chart group by time-tag (week, month or year season) and georeferencing column (city, country, region).
7. The user is able to, in a “tabbed-window”, create a new workspace.
8. In this new workspace, the UI receives “allergic diseases” and “<country name X>”.
9. The system searches the database and connected systems and presents the user with a list of matching datasets.
10. The user selects the dataset it is interested in, the system checks whether access to the dataset is allowed to signed in user, if does, the system shows possible interaction with the dataset, else the system send a permission request to the owner of the dataset.
11. The user selects the “plot in chart diagram”-interaction, selects the columns the user is interested in, selects the columns which will be used as time-tag (week, month or year season), selects the columns which will be used for georeferencing, select that rows should be grouped by selected columns with same values and that number of rows in a group should be also shown in chart diagram.
12. The user asks the system to plot these data in a bar chart, in the same plot as constructed by the system in 6, and to group data using time-tag and georeferencing columns. The user is then able to derive the relation between environmental data and allergic diseases data by time and geospatial dimensions, based on the bar chart.
13. Besides that, the user uses one of the basic analysis tools in the menu of the bar chart: regression analysis.
14. The system performs a linear regression analysis on the 2 different data sets in the bar chart and plots these in the bar chart visualization.
15. The user asks the system to calculate predicted values of a number of allergic diseases for missing allergic data and plot these values in the bar chart visualization.
16. The user asks the system to visualize the results in the map using the selected georeferencing columns. Allergic diseases can be plotted on a map using different colors for different allergic diseases and period of the year (week, month, season), and size of the circle can represent probability (real or predicted number) of allergic reaction.
17. The user clicks on “Download as JPG” and the system exports the visualization to JPG and uploads it to the user’s client.

Post-conditions

- The user is served with one or multiple windows in which the plots are visualized.
- The visualizations are in a separate window, to ensure flexible integration with other personal workflows of the user.
- The user is able to view the underlying raw dataset.



VUC 9: Impact of lifestyle on health

Disciplines

Health, psychology, cognitive science, science of populations (Demographic)

Short case explanation

The <researcher> travel the VRE4EIC environment, take a look on his notification and manage them. After a regard in the funding calls menu, he/she notices that one of them could match with a colleague's project. The <researcher> takes opportunity to forward it through the communication parts of VRE4EIC.

The <researcher> import data that he/she has in possession and research data from a certain country (region). The <researcher> searches for different datasets containing dates, age of persons, blood pressure, heart rate, lung or liver cancer for the certain country (region). Subsequently, the <researcher> is looking for a dataset based on legal drugs addiction like nicotine and alcohol and also on diet for the certain country (region).

After regrouping all data, he/she is able to study relations between people's lifestyle and disease. The <researcher> is able to select rows or columns in both datasets for a more specifics approach. The <researcher> wishes to use VRE4EIC's services to proceed and run some mathematical analysis for a better analysis. In order to get a more comprehensible and visual result, the <researcher> plot the data in a line graph. The data visualization represents a quick overview on possible correlation between lifestyle and chronic diseases.

Furthermore, this investigation can be used for tobacco, alcohol prevention and to warn people on chronic disease by means of diet and lifestyle changes. The majority of the underlying cause relates to modifiable and preventable lifestyle choices made by peoples.

Use case description

As a <health researcher>, I want to study different <people's lifestyles on smoking, drinking and diet> with the aim to <find some correlations with certain chronic disease> in order to publish these relations, to prevent and create awareness of risks on certain lifestyles.

Actors involved in the user case

System user

- Researcher

Priority

Medium

Pre-conditions

The user has to be logged in.

Flow of events – user view

1. <researcher> reorganize notifications.
2. <researcher> use communications services.
3. <researcher> import his data.
4. <researcher> chooses country or region of interest.
5. <researcher> request a medical dataset.
6. <researcher> request a drugs addiction dataset.
7. <researcher> 'travel', observe the datasets.
8. <researcher> performs mathematical analysis with the services provide.
9. <researcher> plots the data points of medical data in a line graph.
10. <researcher> plots the data points of addiction data in the same line graph.
11. <researcher> performs an analysis based on the visualization.
12. <researcher> uploads results.
13. <researcher> make conclusion on the study.
14. <researcher> share investigation and results.

System workflow – system view

1. The UI (User Interface) receive the request to access to <notifications>.

2. The system connects to notifications, display newest notification received for the users log.
3. The UI receives the action to enter into <funding calls> section.
4. The system connects to the funding calls database and show what the user have filtered.
5. The user order to the system to forward the selected overview calls to another user of the system.
6. The UI receives the input "medical data" and "country name".
7. The system searches the database matching the input criteria and presents a list of matching datasets to the user via the UI.
8. The UI receives the selected dataset entry and requests the backend to fetch the entire dataset from the database and loads this onto the client.
9. The user selects the dataset he/she is interested in, the system checks whether access to the dataset is allowed to the user which is signed in, if does, the system shows possible interaction with the dataset, else the system send a dataset permission request to the owner of the dataset.
10. The UI receives the input to visualize the columns "disease", "type of cancer", "age" and "dates".
11. The user is able to, in a "tabbed-window", create a new workspace in which the user searches for a dataset regarding legal drugs information.
12. In this new workspace, the UI receives the input "legal drugs addiction" and "country name".
13. The system searches the database matching the input criteria and presents a list of matching datasets to the user via the UI.
14. The UI receives the selected dataset entry and requests the backend to fetch the entire dataset from the database and loads this onto the client.
15. The UI receives the input to visualize the columns "alcohol", "nicotine", "age", "diet" and "dates".
16. The user selects the "mathematical analysis"-interaction, in order to correct the outlier's values, and apply average calculation on the sample of population.
17. The user selects the "graph line"-interaction, selects the columns he/she is interested in, and asks the system to plot these data in a graph line where both datasets request in 6 and 12 are represents.
18. Besides that, the user uses one of the basic analysis tools in the menu of the graph line: regression analysis.
19. The system performs a linear regression analysis on the 2 different data sets and plots these in the visualization.
20. The user select in the UI to let the analysis and results discoverable and reusable with other users of the system
21. Based on the observations, the user downloads the results for offline use, the system exports the visualization to an image formats (JPG, PNG, …).

Post-conditions

- The user is served with one or multiple windows in which the plots are visualized.
- The visualizations are in a separate window, to ensure flexible integration with other personal workflows of the user.
- The user is able to view the underlying raw dataset.

VUC 10: Urban development in Netherlands

Disciplines

Demographic, transportation, mobility, social sciences

Short case explanation

An <urban planner> wants to do research on the different public transport options in a city and how they are used in order to increase the mobility of people living in the city. The study should represent correlation between demographic development of population and their travel. Dataset with development of populations in Netherlands and transportation data are necessary for investigate on the subject.

The <urban planner> searches for a dataset containing the locations, dates and times, age of persons, numbers of trips, status classification (active or inactive, with or without car, students, preschoolers), for a certain city in Netherlands. Subsequently, the <urban planner> searches for a dataset containing the transportation, type of transport, traffic flow, travel pattern and roadworks for the city.

Based on these datasets, the <urban planner> is able to find relations and correlations between demographic development and transportation data by using methods of analysis, data collection and methods of mathematical statistics. The <urban planner> is able to select rows and columns in all datasets and plot the data in a bar chart so that relevant data are easy to compare.

This study can be used to predict the sustainability and efficiency of news or emergent public transport options, to reduce traffic jam and to help this flow of people to travel: safer, faster and eco-friendlier. Furthermore, the prediction can be used for better organization of transportation companies and to help <urban planner> elaborates news plans for road construction or dedicate road for public transport only. Also, if the mobility of people living the city is not increasing, the <urban planner> should analyse whether the public transport infrastructure or the travel pattern should be better rethink and also consider to implement news transport options.

Use case description

As a <city planner> I would like to be able to <access data from the public transport cards in the Netherlands> so that I can gain an <understanding of how people move through the city> in order to optimize the public transport in the city.

Actors involved in the user case

System user

- Urban planner

Priority

Medium

Pre-conditions

The user has to be logged in.

Flow of events – user view

1. <urban planner> chooses country or region of interest.
2. <urban planner> request a demographic dataset.
3. <urban planner> plots the data points of demographic data in a bar chart.
4. <urban planner> request a transportation dataset.
5. <urban planner> plots the data points of transportation data in a bar chart plot.
6. <urban planner> request a traffic flow dataset.
7. <urban planner> plots the data points of traffic data in a bar chart plot.
8. <urban planner> performs an analysis based on the visualization.
9. <urban planner> performs an analysis based methods of mathematical statistics.
10. <urban planner> creates a model which represent the correlation between demographic data, traffic data and transport time.
11. <urban planner> uploads results.
12. <urban planner> share investigation and results.

13. <urban planner> plots the data points of weather data in a bar chart plot with other users of the system.

System workflow – system view

1. The UI receives the input "demographic data" and "country name".
2. The system searches the database matching the input criteria and presents a list of matching datasets to the user via the UI.
3. The UI receives the selected dataset entry and requests the backend to fetch the entire dataset from the database and loads this onto the client.
4. The user selects the dataset he/she is interested in, the system checks whether access to the dataset is allowed to the user which is signed in, if does, the system shows possible interaction with the dataset, else the system send a dataset permission request to the owner of the dataset.
5. The UI receives the input to visualize the columns "locations", "number of trips", "status classification", "age" and "dates".
6. The user is able to, in a "tabbed-window", create a new workspace in which the user searches for a dataset regarding transportation information.
7. In this new workspace, the UI receives the input "transportation" and "country name".
8. The system searches the database matching the input criteria and presents a list of matching datasets to the user via the UI.
9. The UI receives the selected dataset entry and requests the backend to fetch the entire dataset from the database and loads this onto the client.
10. The UI receives the input to visualize the columns "traffic flow", "roadwork", "travel pattern", "type of travel" and "dates".
11. The user selects the "mathematical analysis"-interaction, in order to make statistics on dataset request in 1 and 7.
12. The user selects the "graph line"-interaction, selects the columns he/she is interested in, and asks the system to plot these data in a graph line where both datasets request in 1 and 7 are represents.
13. The user select in the UI to let the analysis and results discoverable and reusable with other users of the system
14. Based on the observations, the user downloads the results for offline use, the system exports the visualization to an image formats (JPG, PNG, …).

Post-conditions

- The user is served with one or multiple windows in which the plots are visualized.
- The visualizations are in a separate window, to ensure flexible integration with other personal workflows of the user.
- The user is able to view the underlying raw dataset.

VUC 11: A.I. to diagnose mental health issues

Disciplines

Healthcare, A.I., Neuroscience

Short case explanation

A <neuroscientist> wants to know whether it would be possible to develop an AI that could diagnose and help people with mental health issues. Understanding this relation, contribute to improve artificial intelligence (A.I.) techniques to cater a safe and affordable treatment for mental patient. The basic aim is to find the technique which is most accurate to diagnose the mental health issues.

The <neuroscientist> is looking for a dataset containing medical data on three basic mental health problems like anxiety disorders, depression and attention deficit disorder in a certain region. This dataset contains information on symptoms, age of patient, birth condition and date. Besides, the <neuroscientist> has identified different machine learning techniques that can be applied for analysing patient. The different techniques, resume in a dataset, are compare regarding their performances on several scales of accuracy on selected attributes like ROC (Receiver Operating Characteristic) area values, accuracy measure or kappa statistics.

Based on this information the <neuroscientist> is able to find relations between mental health issues and the better machine learning techniques to diagnose mental problems. Considering the very minimal dataset, the investigation was difficult to complete. The <neuroscientist> decide to update data with his/her results and to share it with other users in hope to keep evolving the dataset. Thereafter, the <neuroscientist> is able to export the used data in csv/ms excel format.

This study can be used to determinate several options of suitable sets of treatments for a specific patient based on the patient's health condition. This investigation can also be reused to keep an eye on financial matters, that everyone get access to the treatments indifferently where they lived.

Use case description

As a <neuroscientist> I want to <test people with mental health issues> so that I can <gain insight in patterns shown by people with mental health issues> in order to develop an AI for diagnose and help these people.

Actors involved in the user case

System user

- Neuroscientist

Priority

Medium

Pre-conditions

The user has to be logged in.

Flow of events – user view

1. <neuroscientist> chooses country or region of interest.
2. <neuroscientist> request a medical dataset.
3. <neuroscientist> plots the data points of symptoms data in a bar chart plot
4. <neuroscientist> request a machine learning dataset.
5. <neuroscientist> plots the data points of accuracy measure data in same bar chart plot.
6. <neuroscientist> performs an analysis based on the visualization.
7. <neuroscientist> creates a model which represent the correlation between medical data and machine learning data.
8. <neuroscientist> export dataset results.
9. <neuroscientist> share investigation and results.

System workflow – system view

1. The UI receives the input "medical data" and "country name".
2. The system searches the database matching the input criteria and presents a list of matching datasets to the user via the UI.
3. The UI receives the selected dataset entry and requests the backend to fetch the entire dataset from the database and loads this onto the client.
4. The user selects the dataset he/she is interested in, the system checks whether access to the dataset is allowed to the user which is signed in, if does, the system shows possible interaction with the dataset, else the system send a dataset permission request to the owner of the dataset.

5. The UI receives the input to visualize the columns “symptoms”, “age of patient”, “birth condition” and “dates”.
6. The user is able to, in a “tabbed-window”, create a new workspace in which the user searches for a dataset regarding machine learning information.
7. In this new workspace, the UI receives the input “machine learning” and “country name”.
8. The system searches the database matching the input criteria and presents a list of matching datasets to the user via the UI.
9. The UI receives the selected dataset entry and requests the backend to fetch the entire dataset from the database and loads this onto the client.
10. The UI receives the input to visualize the columns “ROC area value”, “accuracy measure”, “kappa statistics” and “dates”.
11. The user selects the “bar chart”-interaction, selects the columns he/she is interested in, and asks the system to plot these data in a bar chart where both datasets request in 1 and 7 are represents.
12. The user select in the UI the shareable button, to let the analysis and results discoverable and reusable with other users of the system
13. Based on the observations, the user downloads the results for offline use, the system exports the database in csv/ms excel files format.

Post-conditions

- The user is served with one or multiple windows in which the plots are visualized.
- The visualizations are in a separate window, to ensure flexible integration with other personal workflows of the user.
- The user is able to view the underlying raw dataset.

VUC 12: Locate DDoS attacks

Disciplines

Security, geolocation

Short case explanation

A <researcher> wants to detect DDoS attacks and where they are located. One of the important things is to be able to detect those attacks directly at their source computers or the first kilometre routers. This strategy brings benefits such as small collateral damage, easy trace back or congestion avoidance. The investigation should represent correlation between traffic flow DDoS attacks and their localisation. A dataset representing type of DDoS attacks and location where they start are required to deal on this subject.

A <researcher> searches for a dataset containing access log for some web server. The <researcher> detects IP addresses of potential botnet members. Afterwards, the <researcher> searches for a dataset with correlation between IP addresses and geolocation such as GeoLite database.

Based on these datasets the <researcher> is able to find relations between potential botnet members and its location. After that, the researcher analyzes weak and strong security regions. Weak security regions can take into consideration application of security practices from strong security regions.

Use case description

As a <researcher> I want to <be able to analyse the traffic flow of DDoS amplification attacks> so that I can <gain insight in the security practices in different regions> in order to develop security governance practices in weak areas.

Actors involved in the user case

System user

- Researcher

Priority

Medium.

Pre-conditions

The user has to be logged in.

Flow of events – user view

1. <researcher> requests a server access log dataset.
2. <researcher> requests a IP address geolocation dataset.
3. <researcher> requests merging of datasets by IP address column.
4. <researcher> invokes plotting of the merged information on an Open Street Maps visualization
5. <researcher> analyses strong and weak regions taking into account the number of access request from various regions (high number of access request represents potential botnet members for DDoS attack)

System workflow – system view

1. The UI receives the input “access log”
2. The system connects to the database and connected systems to request datasets containing “access log” in the title or metadata of the dataset.
3. The system presents, via the UI, a list of datasets it retrieved from the database to the user.
4. The user selects the dataset it is interested in, the system checks whether access to the dataset is allowed to signed in user, if does, the system shows possible interaction with the dataset, else the system send a permission request to the owner of the dataset.
5. The user is able to, in a “tabbed-window”, create a new workspace. In this new workspace, the UI receives “IP geolocation”.
6. The system searches the database and connected systems and presents the user with a list of matching datasets.
7. The user selects the dataset it is interested in, the system checks whether access to the dataset is allowed to signed in user, if does, the system shows possible interaction with the dataset, else the system send a permission request to the owner of the dataset.
8. The user selects the “merge data”-interaction, selects the columns from both datasets the user is interested in, selects the columns which will be used for merging data (IP address), selects the columns which will be used for georeferencing.
9. The user asks the system to visualize the results in the map using the selected georeferencing columns. Request to the server access can be plotted on a map using different size representing number of access.
10. The user clicks on “Download as JPG” and the system exports the visualization to JPG and uploads it to the user’s client.

Post-conditions

- The user is served with window in which the data are visualized.
- The visualizations are in a separate window, to ensure flexible integration with other personal workflows of the user.
- The user is able to view the underlying raw dataset.

VUC 13: Seismic activity and global warming

Disciplines

Seismology, geology, climatology

Short case explanation

This visionary use case describes a <researcher> study showing that increasing seismic activity for the globe's high geothermal flux areas, an indicator of increasing geothermal forcing, is highly correlated with average globe temperatures from 1979 to 2015. Temperatures have caused snow and ice extent to decline, sea levels to rise and also amplified extreme events in the atmosphere including tropical cyclone intensity, heavy precipitation episodes and hot spells.

To test the hypothesis that increasing geothermal flux has contributed to recent global warning, the researcher would like to obtain data from different fields.

The researcher place earth's geothermal heat source in a heat map visualization so he is able to show the locations of the geothermal heat zones. More intense means higher geothermal flux. Numerous studies have shown that the impact of the total amount of heat from geothermal sources are significant. The underwater geothermal heat can amplify the transfer of heat from the ocean to the overlying atmosphere.

The researcher also search data on frequency and impact of seismic activity. The data can be plotted on a map using the Open Street Maps API to easily localised them. For the experiment, seismic activity is used as a proxy for geothermal flux. Seismicity is strongly associated with geothermal flow and it is easily deduced that increasing seismicity will indicate increasing geothermal flux. From the earthquake datasets, the researcher select only input events with magnitudes between 4.0 and 6.0 (MAG4/6). Low-magnitude events are excluded from the analysis since only earthquakes with magnitudes greater than 3.9 are detected at sufficiently high levels.

Using this Open Street Maps visualisation of geothermal sources combining with plot of earthquakes, the researcher is able to locate certain regions to analysed and provide a correlation between the localisation of higher earthquake and biggest sources of geothermal heat flux. In a second time the researcher want to compare dataset between seismic frequencies in the more active areas for specific periods with the anomalies found in temperature records. By comparing a bar chart and a line chart the researcher is able to see if these areas showed any changes before or during a specific period, that explains the variation in the earth's surface temperature. Conversely, if seismic activity gaps downward, this may indicate the onset of a cooling period.

Based on the information regarding the temperature anomalies, the researcher searches for datasets to quantifying the anomaly. Many global temperature, surface-based datasets contain errors and require corrections and/or adjustments in order to be truly representative. The urban contamination, deforestation and increased cultivation, introduce warning biases into many station records.

Use-case description

As a <researcher> I want to be able to < map and deduced that increasing seismicity will indicate increasing geothermal flux and to compare the earthquake activities with the temperature since the past year> so that I can <gain an understanding of how geothermal heat activities on earth contributes to rise the average globe temperatures> in order to warn the climate community to explore the impacts of geothermal flux.

Actors involved in the user case

System user

- Researcher

Priority

Medium

Pre-conditions

The user has to be logged in.

Flow of events – user view

Basic sequence and needed steps

1. Obtain data regarding the earth's geothermal heat flux.
2. Locate the source of geothermal heat flux in a heat map.
3. Obtain data on frequency and impact of seismic activity.
4. Plot the earthquake information on an Open Street Maps visualization.
5. Based on the high magnitude earthquake position in the map, a proximity correlation with geothermal heat flux can be made.
6. Based on the earthquake activities, the researcher is able to deduce by comparing with the anomalies found in temperature records the variation on the earth's surface temperature.
7. The researcher is able to download the results or share the results with other users of the system.

System workflow – system view

1. The UI (User Interface) receives the input: "earthquake", "geothermal heat flux"
2. Backend system connects to the database and searches for datasets that have those input parameters in its title.
3. The backend retrieves the datasets matching the input criteria and represents this list of possibilities to the user.
4. The UI receives the selected dataset entry and requests the back-end to fetch the entire dataset from the database and loads this onto the client.
5. The UI receives the action to visualize the columns "earthquake name", "coordinates", "magnitude" and "geothermal heat flux source"
6. The system recognizes coordinate format, so it retrieves the corresponding area from the Open Street Maps API.
7. Open Street Maps canvas is pained using the data points from the input columns, which results in a heat map.
8. The painted heat map, without sacrificing Open Street Maps features and details, is served to the user. Earthquakes should thus be still visible in the plot.
9. The user is able to, in a "tabbed-window", create a new workspace in which the user searches for a dataset regarding the globe's surface-based temperature information.
10. The user selects a bar chart graph on earthquake magnitude and the system retrieves the required data from the underlying database system.
11. The user selects a line chart graph on surface-based temperature and the system retrieves the required data from the underlying database system.
12. Based on the observations, the user downloads the results for offline use. In addition, the user is able to share the results with other users of the system.

Post-conditions

- The user is served with one or multiple windows in which the plots are visualized.
- The visualizations are in a separate window, to ensure flexible integration with other personal workflows of the user.
- The user is served with one or multiple windows in which charts are display.
- The user is able to view the underlying raw dataset.

13.2 Annex B: List of achieved requirements

The following table presents a list of the requirements and the functions (FUN) they are linked to. By looking at the FUN that have been realised within the project, we can see the number of requirements that are completed.

RQ	UC	FUN	FUN (Done)
IRQ5	UC001	**Fun1 (Simple Search)**	X
PRQ10	UC020	**Fun1 (Simple Search)**	X
PRQ11	UC020	**Fun1 (Simple Search)**	X
PRQ14	UC020	**Fun1 (Simple Search)**	X
PRQ15	UC020	**Fun1 (Simple Search)**	X
CLRQ6	UC008	**Fun10 (Instrument Monitoring)**	
CLRQ7	UC006/UC009	**Fun10 (Instrument Monitoring)**	
CLRQ8	UC009	**Fun10 (Instrument Monitoring)**	
CLRQ11	UC008	**Fun11 (Data Collection)**	X
CLRQ12	UC008	**Fun11 (Data Collection)**	
CLRQ13		**Fun12 (Data Sampling)**	
CLRQ14	UC003/UC005	**Fun12 (Data Sampling)**	
CLRQ15	UC002/UC013	**Fun12 (Data Sampling)**	
CLRQ16	UC013	**Fun12 (Data Sampling)**	
CRQ1	UC003	**Fun12 (Data Sampling)**	
CRQ7	UC002	**Fun12 (Data Sampling)**	
CRQ8	UC002	**Fun12 (Data Sampling)**	
CTRQ9	UC003/UC015/UC037	**Fun12 (Data Sampling)**	
IRQ1	UC001/UC014	**Fun12 (Data Sampling)**	
IRQ3	UC001/UC020	**Fun12 (Data Sampling)**	
IRQ4	UC001	**Fun12 (Data Sampling)**	
PRQ29	UC004/UC007/UC027	**Fun12 (Data Sampling)**	
PRQ35		**Fun12 (Data Sampling)**	
PRQ7	UC015/UC021	**Fun12 (Data Sampling)**	
PVRQ1	UC031	**Fun12 (Data Sampling)**	
PVRQ2	UC032	**Fun12 (Data Sampling)**	

PVRQ3	UC032	**Fun12 (Data Sampling)**	
PVRQ4	UC015/UC031/UC050/UC051	**Fun12 (Data Sampling)**	
CLRQ17	UC003/UC010/UC014	**Fun13 (Resource Registration)**	X
CLRQ18	UC015	**Fun13 (Resource Registration)**	X
CLRQ9	UC008	**Fun13 (Resource Registration)**	X
CRQ2	UC003/UC005	**Fun13 (Resource Registration)**	X
CRQ3	UC003/UC005	**Fun13 (Resource Registration)**	X
CRQ6	UC010	**Fun13 (Resource Registration)**	X
CTRQ15	UC040/UC041/UC047/UC054	**Fun13 (Resource Registration)**	X
IRQ2	UC001/UC014	**Fun13 (Resource Registration)**	X
IRQ4	UC001	**Fun13 (Resource Registration)**	X
PRQ4	UC010/UC016/UC017	**Fun13 (Resource Registration)**	X
PRQ5	UC003/UC006/UC010/UC016/UC017	**Fun13 (Resource Registration)**	X
PRQ6	UC003/UC013/UC018	**Fun13 (Resource Registration)**	X
PRQ8	UC019	**Fun13 (Resource Registration)**	X
CRQ4	UC003/UC006/UC010	**Fun14 (Resource Update)**	X
CRQ5	UC004	**Fun15 (Workflow Enactment)**	
ORQ2	UC003/UC008	**Fun15 (Workflow Enactment)**	X
ORQ4	UC003/UC008	**Fun15 (Workflow Enactment)**	X
PRQ26	UC004/UC007	**Fun15 (Workflow Enactment)**	
PRQ28	UC004/UC007/UC027	**Fun15 (Workflow Enactment)**	X
CRQ6	UC010	**Fun16 (Access Control)**	
PRQ1	UC016	**Fun17 (Resources Annotation)**	X
PRQ2	UC016	**Fun17 (Resources Annotation)**	X
PRQ32	UC058	**Fun17 (Resources Annotation)**	X
PRQ34	UC030	**Fun17 (Resources Annotation)**	X
PRQ3	UC017	**Fun18 (Metadata Harvesting)**	X
PRQ31	UC020	**Fun19 (Data Discovery)**	X
PRQ9	UC020	**Fun19 (Data Discovery)**	X
IRQ5	UC001	**Fun2 (Cross Search)**	X

PRQ12	UC020	**Fun2 (Cross Search)**	X
PRQ14	UC020	**Fun2 (Cross Search)**	X
PRQ15	UC020	**Fun2 (Cross Search)**	X
PRQ20	UC020	**Fun2 (Cross Search)**	X
CTRQ4	UC034	**Fun20 (User/Agent Registration)**	X
CTRQ5	UC035	**Fun20 (User/Agent Registration)**	X
CTRQ7	UC034	**Fun20 (User/Agent Registration)**	X
CTRQ1	UC028/UC034	**Fun21 (Agent authentication)**	X
CTRQ31	UC049	**Fun21 (Agent authentication)**	X
CTRQ2	UC033	**Fun22 (Continuous Access)**	X
CTRQ10	UC015/UC038	**Fun23 (Update Alert)**	X
CTRQ8	UC015/UC038	**Fun23 (Update Alert)**	X
CTRQ28	UC003	**Fun24 (Resource Connection)**	
IRQ5	UC001	**Fun3 (Advanced Search)**	X
PRQ13	UC020	**Fun3 (Advanced Search)**	X
PRQ14	UC020	**Fun3 (Advanced Search)**	X
PRQ15	UC020	**Fun3 (Advanced Search)**	X
PRQ16	UC020	**Fun3 (Advanced Search)**	X
PRQ17	UC022	**Fun4 (Data viewing)**	X
PRQ18	UC022	**Fun5 (Dataset Preselection)**	X
PRQ19	UC022	**Fun6 (Dataset Customisation**	X
CLRQ1	UC007	**Fun7 (Instrument Integration)**	
CLRQ10	UC007	**Fun8 (Instrument Configuration)**	
CLRQ2	UC007	**Fun8 (Instrument Configuration)**	
CLRQ4	UC007/UC008	**Fun8 (Instrument Configuration)**	
CLRQ5	UC008	**Fun8 (Instrument Configuration)**	
CLRQ3	UC007	**Fun9 (Instrument calibration)**	
CLRQ4	UC007/UC008	**Fun9 (Instrument calibration)**	

13.3 Annex C: Completion level of use-cases

Based on the table presented in appendix B, the following table indicates, for each use-case, how many requirements it includes, as well as the number of them that are already implemented. This allows to indicate some kind of indicator about the level of completion of the use-cases.

Use cases	**Requirements per use case**	**Requirements implemented**	**Complete**
UC001	3	3	100.00%
UC002	2	0	0.00%
UC003	17	8	47.00%
UC004	7	1	14.00%
UC005	2	2	100.00%
UC006	3	3	100.00%
UC007	8	1	13.00%
UC008	11	4	36.00%
UC009	3	0	0.00%
UC010	5	5	100.00%
UC011	1	0	0.00%
UC012	1	0	0.00%
UC013	1	1	100.00%
UC014	2	2	100.00%
UC015	4	4	100.00%
UC016	2	2	100.00%
UC017	3	3	100.00%
UC018	2	1	50.00%
UC019	3	1	33.00%
UC020	15	13	87.00%
UC021	0	0	0.00%
UC022	3	3	100.00%
UC023	5	0	0.00%
UC024	1	0	0.00%
UC025	1	0	0.00%
UC026	1	0	0.00%
UC027	1	1	100.00%

UC028	2	1	50.00%
UC029	1	0	0.00%
UC030	1	1	100.00%
UC031	2	0	0.00%
UC032	0	0	0.00%
UC033	3	1	33.00%
UC034	4	3	75.00%
UC035	3	1	33.00%
UC036	1	0	0.00%
UC037	0	0	0.00%
UC038	6	2	33.00%
UC039	3	0	0.00%
UC040	9	1	11.00%
UC041	5	1	20.00%
UC042	1	0	0.00%
UC043	3	0	0.00%
UC044	2	0	0.00%
UC045	6	0	0.00%
UC046	1	0	0.00%
UC047	12	1	8.00%
UC048	1	0	0.00%
UC049	1	1	100.00%
UC050	3	0	0.00%
UC051	8	0	0.00%
UC052	2	0	0.00%
UC053	1	0	0.00%
UC054	5	1	20.00%
UC055	1	0	0.00%
UC056	1	0	0.00%
UC057	2	0	0.00%

UC058	1	1	100.00%
UC059	2	0	0.00%