Large-scale EXecution for Industry & Society

Deliverable D6.1

Baseline Scenarios and Requirements

Co-funded by the Horizon 2020 Framework Programme of the European Union

Grant Agreement Number 825532

ICT-11-2018-2019 (IA - Innovation Action)

DELIVERABLE ID | TITLE	D6.1 | Baseline scenarios and requirements
RESPONSIBLE AUTHOR	Danijel Schorlemmer (GFZ)
WORKPACKAGE ID | TITLE	WP6 | Earthquake and Tsunami large scale pilot
WORKPACKAGE LEADER	CEA
DATE OF DELIVERY (CONTRACTUAL)	30/06/2019 (M06)
DATE OF DELIVERY (SUBMITTED)	24/06/2019 (M06)
VERSION | STATUS	V1.1 | Final
TYPE OF DELIVERABLE	O (Other)
DISSEMINATION LEVEL	PU (Public)
AUTHORS (PARTNER)	Thierry Goubier (CEA), Danijel Schorlemmer (GFZ), Thomas Beutin (GFZ), Natalja Rakowski (AWI), Sven Harig (AWI)
INTERNAL REVIEW	Tomas Martinovic (IT4I), Seán Murphy (CYC)

Project Coordinator: Dr. Jan Martinovič – IT4Innovations, VSB – Technical University of Ostrava

E-mail: jan.martinovic@vsb.cz, **Phone:** +420 597 329 598, **Web:** https://lexis-project.eu

DOCUMENT VERSION

VERSION	MODIFICATION(S)	DATE	AUTHOR(S)
0.1	Initial template	01/06/2019	Thierry Goubier (CEA)
0.2	GFZ Benchmark description	08/06/2019	Danijel Schorlemmer (GFZ); Thomas Beutin (GFZ)
0.3	Workflow description	15/06/2019	Thierry Goubier (CEA)
0.4	Tsunami Benchmark description	18/06/2019	Natalja Rakowski (AWI)
0.5	Introduction, executive summary	20/06/2019	Thierry Goubier (CEA)
0.6	Polishing, Consolidation	22/06/2019	Danijel Schorlemmer (GFZ)
0.7	Review of the document	25/06/2019	Tomáš Martinovič (IT4I)
0.8	Review of the document	26/06/2019	Seán Murphy (CYC)
1.0	Revision after internal reviews	28/06/2019	Thierry Goubier (CEA); Danijel Schorlemmer (GFZ)
1.1	Final review by the coordinator	30/06/2019	Kateřina Slaninová (IT4I); Ekaterina Grakova (IT4I); Jan Martinovič (IT4I)

GLOSSARY

ACRONYM	DESCRIPTION
GIS	Geographical Information System: a database and additional functions for data with a geographical meaning and coordinates
SEM	Satellite-based Emergency Mapping: the process of using remote sensing data to produce maps for emergencies
BPMN	Business Process Model Notation: a graphical notation to describe and model business processes
WP6	Work Package 6
DEM	Digital Elevation Model

TABLE OF PARTNERS

ACRONYM	PARTNER
Avio Aero	GE AVIO SRL
AWI	ALFRED WEGENER INSTITUT HELMHOLTZ ZENTRUM FUR POLAR UND MEERESFORSCHUNG
BLABS	BAYNCORE LABS LIMITED
Bull/Atos	BULL SAS
CEA	COMMISSARIAT A L ENERGIE ATOMIQUE ET AUX ENERGIES ALTERNATIVES
CIMA	Centro Internazionale in Monitoraggio Ambientale - Fondazione CIMA
CYC	CYCLOPS LABS GMBH
ECMWF	EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS
GFZ	HELMHOLTZ ZENTRUM POTSDAM DEUTSCHESGEOFORSCHUNGSZENTRUM GFZ
IT4I	VYSOKA SKOLA BANSKA - TECHNICKA UNIVERZITA OSTRAVA / IT4Innovations National Supercomputing Centre
ITHACA	ASSOCIAZIONE ITHACA
LINKS	FONDAZIONE LINKS / ISTITUTO SUPERIORE MARIO BOELLA ISMB
LRZ	BAYERISCHE AKADEMIE DER WISSENSCHAFTEN / Leibniz Rechenzentrum der BAdW
NUM	NUMTECH
O24	OUTPOST 24 FRANCE
TESEO	TESEO SPA TECNOLOGIE E SISTEMI ELETTRONICI ED OTTICI

LIST OF FIGURES

TABLE OF CONTENTS

EXECUTIVE SUMMARY

This deliverable describes the processes, performance benchmarks and associated baselines relating to the earthquake and tsunami large scale pilot of the LEXIS project. The project objectives are to implement and demonstrate a federation of resources and technologies at the convergence of big data and high-performance computing, around three pillars (infrastructure, orchestration and data management), and this pilot is to demonstrate the effectiveness of the project with a challenging use case.

Position of the deliverable in the whole project context

The LEXIS project relies on three pilots to validate and deploy its technology and infrastructure advances, giving to each pilot its own work package. The Work Package 6 (WP6) is dedicated to the earthquake and tsunami large scale pilot. It starts with a task T6.1, "Define baseline scenarios and requirements", and this deliverable D6.1 is the result of this activity. It specifies in detail the pilot, as it has converged between the partners involved in the WP6 since the project start and establish the framework for the subsequent tasks.

Description of the deliverable

This deliverable contains the description of the tsunami and earthquake large scale pilot workflow, that is the set of computational and human tasks with their dependencies and their interaction with outside systems and events, the targeted scenario the pilot will be tested and demonstrated with, and the description of two benchmarks defining the key performance points that the pilot will use to measure gains, with a technical appendix describing the underlying technologies of the application partners. It demonstrates the complexity of the workflow, the main performance issues identified so far, and the work already under way to succeed in integrating and coordinating all the elements of the pilot.

1 INTRODUCTION

The LEXIS project is developing solutions for federated HPC and Cloud resources, enhanced by the burst buffer technology. This is comprised of three pillars, that is infrastructure, orchestration and data management. The project relies on three pilots representing different domains and exercising the LEXIS platform.

The objective of the large-scale earthquake and tsunami pilot is the implementation of a time-constrained, disaster-related workflow triggered by earthquake events. It combines four main application components: A ShakeMap [1] code from the OpenQuake library [2] to compute ground-motion distributions for earthquakes; TsunAWI [1], a code simulating the propagation and inundation of an earthquake-triggered tsunami; OpenBuildingMap [1], a database of geographical information focusing on the built environment and an associated building classification and loss assessment method allowing detailed damage assessments at the building scale after an earthquake and tsunami event, and a satellite-based emergency mapping process producing post-disaster map products. Those four applications are integrated in a workflow described by the HeScade [1] model of computation, and implemented over the infrastructure, orchestration and data management of the LEXIS platform.

For a proper implementation of the pilot, a reference scenario and baselines are necessary. The scenario will drive the assessment of the pilot implementation by describing the conditions under which the pilot will be demonstrated. Improvements in each component of the workflow are planned beyond the implementation of the integrated flow itself. For that purpose, the two main applications, TsunAWI and OpenBuildingMap, are represented by benchmarks, linked to one or more datasets. Benchmarks and their associated datasets are used internally to assess progress and optimisation relative to the baseline, but also as dissemination assets: we intend to allow actors outside the project partners to work on complementary technologies and experiments to support the project and the implementation ideas. This is reflected in the data management plan (Deliverable D9.2 [3]) where the datasets are described as well.

For a more detailed description of the technical elements of the pilot, we refer to [1], included in annex to this deliverable.

2 WORKFLOW DESCRIPTION

The pilot workflow describes how the different elements interact and their timeline, particularly after an event. A high-level overview of the workflow is listed, followed by a detailed workflow description using the Business Process Modelling Notation (BPMN) graphical language.

2.1 HIGH-LEVEL OVERVIEW

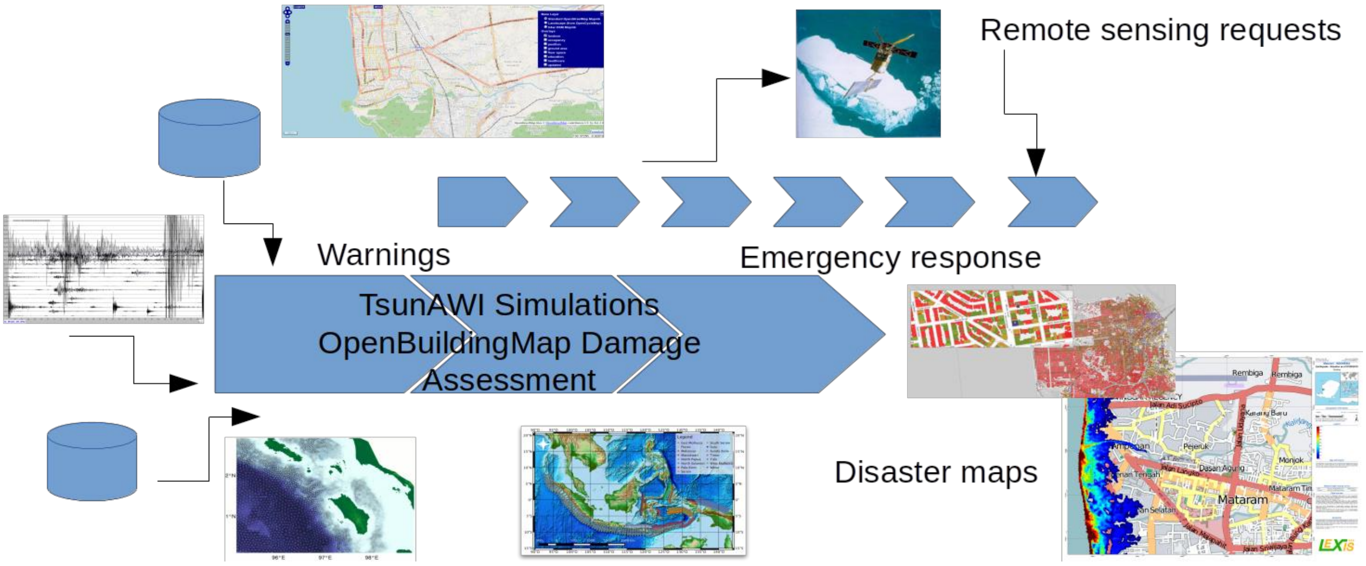

Figure 1 A high-level overview of the earthquake and tsunami workflow

Figure 1 shows the high-level view of the pilot. In the earthquake and tsunami pilot, the OpenBuildingMap from GFZ provides us with worldwide building data, and the TsunAWI tsunami simulation code from AWI provides us with the ability to simulate tsunamis. Upon an earthquake event, OpenQuake will provide a first assessment of ground motions, TsunAWI will simulate a possible tsunami wave and its physical impact, and from OpenBuildingMap loss assessments for the impact of the earthquake and the tsunami will be separately computed. This information will be used to compute emergency mapping products and warnings, to be later completed with remote sensing data.

2.2 DETAILED DESCRIPTION

The goal of the flow (Figure 2) is to provide information at specific time points in an emergency warning system, and in an emergency response process. As such, it is designed to provide information at timely points for emergency warnings, and then more precise information at later points in the emergency response process.

The flow is laid out to match this requirement and improve each of the individual components. A significant challenge is to be able to properly schedule all the flow elements according to the time constraints.

The BPMN notation used in Figure 2 is an informal, yet detailed, notation. We have signals (e.g. “earthquake hypocenter”) triggering processing. Processing is done by tasks, distinguished here by cogs (an application, for e.g. “TsunAWI fast simulation on HPC”), scrolls (a script, for e.g. “Interpolation, raster output”) and a human symbol (done by a human operator, such as “Emergency map production”). Tasks are scheduled according to the arrows connecting them. Each task may have data objects as inputs or outputs, and we distinguish between two types of data objects: files (e.g. “Shakemap”) and databases (e.g. “OpenBuildingMap”). Deadlines for terminating certain tasks are the clock symbols, and a specific task is sending a message to the outside (the “Remote sensing requests” task) and reception is handled by the message symbol “Remote sensing reception”.

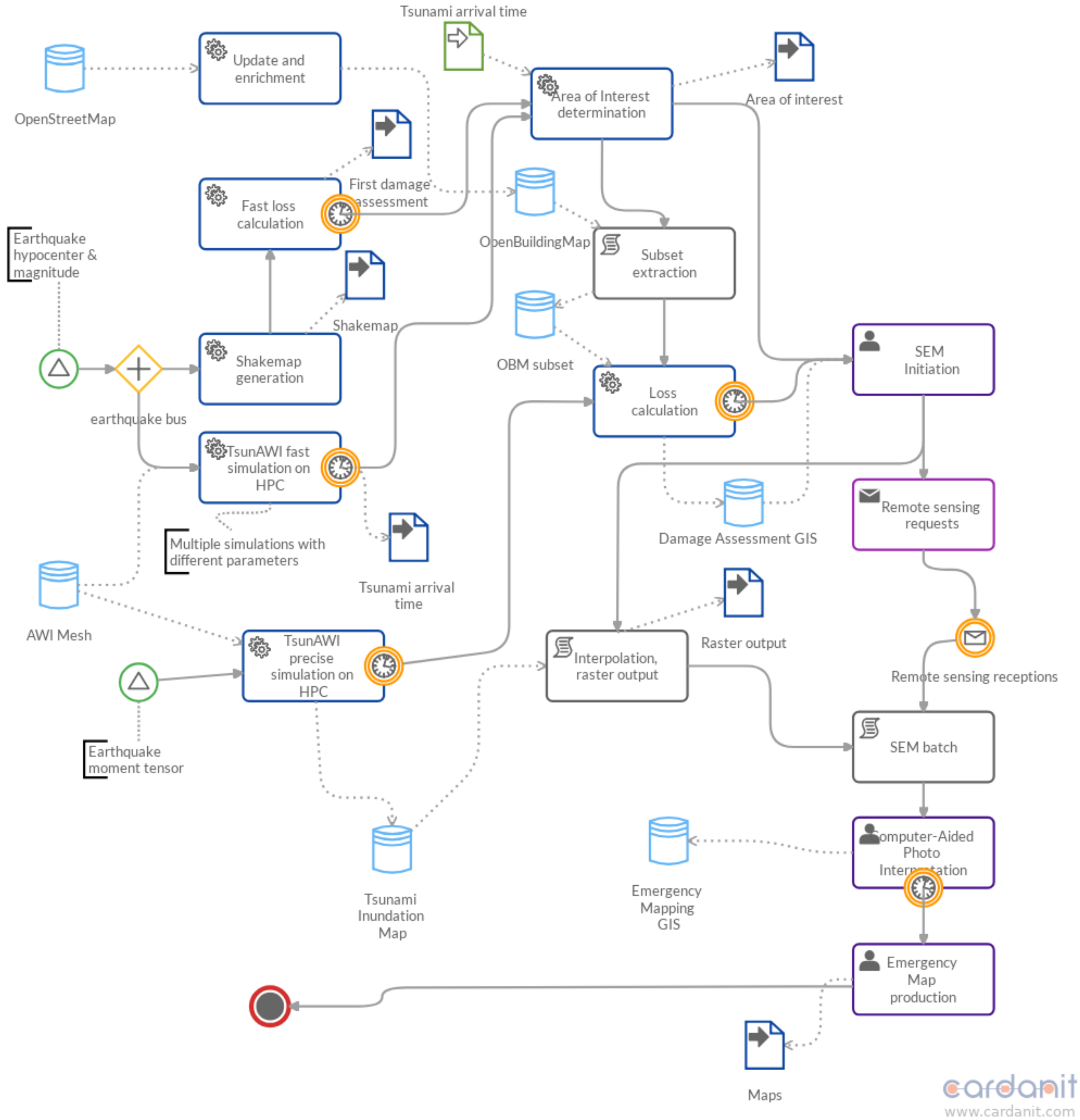

Figure 2 The detailed description of the earthquake and tsunami worklfow

The flow is composed of four groups, according to a coarse grain view of the flow: the OpenStreetMap group, the early earthquake and tsunami group, the precise earthquake and tsunami group, and the Satellite-based Emergency Mapping (SEM) group.

The OpenStreetMap group

This group is online all the time, processing updates from the OpenStreetMap global database every minute, and keeping the OpenBuildingMap up to date.

- **OpenStreetMap**: the OpenStreetMap global database containing state and crowd-sourced data about geographic features of the world (land use, administrative boundaries, roads and paths, waterways, buildings) under constant update and increase (approx. 150,000 new buildings every day).
- **Update and Enrichment**: this process receives updates from the OpenStreetMap database (updated every minute), import those updates in the OpenBuildingMap PostGIS database, and add information to all updated and imported buildings according to a set of predefined rules.
- **OpenBuildingMap**: the global geographical database containing the augmented building information needed for the detailed loss assessment.

The early earthquake and tsunami group

This group is triggered when receiving an earthquake event and has a tight deadline to completion. It produces an estimate of the tsunami arrival time, a shakemap and a fast loss assessment.

- **Earthquake hypocenter and magnitude**: provided by the GEOFON[1] global network of seismometers when an earthquake happens.
- **Shakemap generation**: produces a Shakemap describing the estimated ground shaking and hence the affected area by the shaking.
- **Fast loss calculation**: computes a first damage assessment using a coarse exposure model.
- **TsunAWI fast simulation on HPC**: fast tsunami wave simulation. Produces a first estimate of the tsunami arrival times. Multiple fast simulations can be run at the same time, in parallel, on multiple resources to allow for the exploration of multiple scenarios and hypotheses, and to try to run the best simulation for the amount of time available in the workflow schedule while making sure faster simulations are also run to guarantee a result by the deadline.
- **Area of interest determination**: determining which specific areas should be targeted in the mapping process, based on the importance of the estimated damage (presence of a large number of victims, large material damage, critical infrastructure, etc.).
- **Area of interest**: the geographic boundaries of the area of interest.

The precise earthquake and tsunami group

This group is triggered after receiving the earthquake moment tensor, at around 10 minutes after the initial earthquake event. It produces precise wave height and inundation and precise damage assessments, that are used to update the information available to the SEM group.

- **AWI Mesh**: the triangular, pre-computed mesh of the target area, computed out of a Digital Elevation Model (DEM).
- **Earthquake moment tensor**: the updated and more precise earthquake related information, received ten minutes after the earthquake event itself.
- **TsunAWI precise simulation on HPC**: a precise simulation of the tsunami wave and inundation triggered by an earthquake, producing a detailed inundation map, but not in-time for the emergency warning deadline except for far-field warning. Is dependent on a precise knowledge of the earthquake event, that is the earthquake moment tensor, available 10 minutes after the earthquake itself.
- **Tsunami Inundation Map**: the detailed description of the computed inundation for the tsunami triggered by the earthquake event.
- **OpenBuildingMap subset**: the subset of the global OpenBuildingMap database covering the affected areas (areas of interest).
- **Subset extraction**: the process of extracting the relevant subset of the OpenBuildingMap global database, that is the information related to the area of interest/affected area.
- **Loss calculation**: a per-building computation of the damage probabilities, combining both the local shaking and damage from the inundation as computed by the simulations.
- **Damage assessment GIS**: the geographical database containing the results of the loss calculation, available to establish queries and further analyses.

[1] http://geofon.gfz-potsdam.de; http://doi.org/10.14470/TR560404

The SEM group

This group is activated by the computation of the first area of interest results and is also updated when the precise group has completed. It includes some operator tasks, requiring human intervention, and produces data products (database and maps) for emergency response services.

- **Interpolation, raster output**: the process of extracting rasters at various resolutions out of the mesh result data from the TsunAWI simulation.
- **Raster output**: the rasters produced by the interpolation and raster process, that is raster data that can be correlated to remorse sensing data products.
- **SEM initiation**: the beginning of the Satellite-based Emergency Mapping process, and specially the determination of the remote sensing data products that will be needed. This is an operation performed by a trained operator out of the various information provided by the previous stages of the flow.
- **Remote sensing requests**: The products requested to the remote sensing data providers and needed for production of the emergency mapping products. Those are probably the earliest point where the process has a ground truth of the situation after the event, especially in the case of a large event.
- **SEM batch**: part of the Satellite-based Emergency Mapping process which are scripted over the cloud or on remote compute resources.
- **Computer-aided photo interpretation**: work by an operator to combine the remote sensing data products, the other sources (the damage assessment GIS, the inundation results) and estimate from the post-event remote sensing data the state of the areas of interest, and an estimated damage.
- **Emergency Mapping GIS**: the geographical information system storing the results of the computer aided photo interpretation.
- **Emergency Map production**: the process of producing the emergency maps for the Copernicus European emergency network.
- **Maps**: the emergency maps for the areas of interest

Time constraints are applied to this process in two areas: for emergency warning, the first two deadlines after the event [4], and then deadlines that apply to the emergency response process. Emergency warning deadlines target first estimates and fast damage assessment. A second set of deadlines are associated with emergency response decisions, because they are already too late to be applied to the emergency warnings, except in far-field tsunami warnings.

3 BASELINE SCENARIO AND BENCHMARKS

In this section, we will describe the scenario used to evaluate the pilot, ensuring that, on a test case, the pilot will be able to execute itself according to its objectives. Benchmarks being available to characterize the main components of the pilot workflow, they will be listed and described in the following section, so as to be reproducible.

3.1 BASELINE SCENARIO

The scenario is based on the benchmark defined for TsunAWI (see Section 4.1), that is a well-studied possible event, an earthquake with epicenter located offshore the city of Padang, West Sumatra region, Indonesia, triggering a tsunami. The affected area by the earthquake and the tsunami are the city of Padang and the surrounding inhabited areas, as defined in the region extraction operation of the OpenBuildingMap benchmarks. The time constraints are based on the current InaTEWS deadlines, as summarized by the BMKG state agency in [4], particularly the limits for the emission of a warning and termination of the warning. Those times are, out of the T1 to T7 times listed in [4], T1 < 5 minutes for the first warning, and T4 10-60 minutes for the first observation, and T7 > 120 minutes for the end of threat signal. Additional times are linked to the Satellite-based Emergency Mapping process, where 24 hours delay in data acquisition is avoided [1].

3.2 TSUNAWI

The pilot area for the TsunAWI benchmark is located around Padang in Sumatra, Indonesia. The area is prone to earthquakes and the geographical setting of the city with more than 900,000 inhabitants and large parts only slightly above sea level makes it vulnerable to tsunamis as well.

The benchmark investigates a hypothetical earthquake scenario with an epicenter located offshore Padang (99.72°W/1.66°S) and a magnitude Mw = 8.8 The earthquake generates a tsunami and the initial sea surface elevation is generated by the software Rupture Generator [2]. The simulation of the tsunami is carried out with TsunAWI in a triangular mesh covering the region and resolving the city area of Padang with triangle edge length of approximately 20m. The simulation time is 2 hours and uses a time step of 0.1 seconds. It takes about 25 minutes on 2x18 cores Intel Xeon Broadwell on the Linux cluster Cray CS400 at AWI. Wave propagation and inundation is taken into account by the model, the output contains flow depth values referenced to the topography data available in the simulation.

3.3 OPENBUILDINGMAP

For meaningful benchmarks of the development of OpenBuildingMap, we need to consider the full OpenStreetMap global dataset2 as the baseline scenario. Any processing of building data needs increasingly longer times the larger the database becomes. Therefore, we select the full OpenStreetMap planet file for 7 January 2019 as the benchmark core dataset. To benchmark the updating procedures, we provide in addition to the planet file a set of approx. 230,000 minutely differential files (minutely diffs from OpenStreetMap) to be used to replay the updating procedure and measure the processing duration. These published files contain all changes that were submitted to OpenStreetMap for each minute. For operations that either extract or use a subset of the global data, the areas are chosen to be either the same as the TsunAWI benchmark, or areas known to be a representative. All the data required to run the benchmarks are provided for download: http://data.obm.gfz-potsdam.de/Lexis/ and will be later stored at one of long-term repositories if necessary.

The benchmarks are as follow:

- Extracting a region (e.g. a city or county) from the global database of already computed buildings provided by GFZ. This benchmark is separated into several cases to reflect various user scenarios, i.e. extracting areas with many or with few buildings and using simple (i.e. rectangular bounding boxes) and complex polygons as boundaries. We propose the following areas:
    - San Francisco bay area using a complex polygon.
    - Area of greater Tokyo using a bounding box.
    - City of Cologne using the city boundary (complex polygon).
    - Tohoku area as approximately affected by the 2011 tsunami using a bounding box.
    - Padang area (West Sumatra region, Indonesia) using a bounding box.

[2] https://planet.osm.org/

- The time needed for running an update cycle of several weeks (covered by the aforementioned approx. 230,000 minutely diffs) to the full planet file to update the planet-file database.
- Computing all exposure indicators of all buildings within a bounding box of the planet-file database. We use a bounding box of the greater Berlin metropolitan area for this benchmark. In this benchmark, the exposure computation will run on the global dataset.
- A related benchmark for the computationally more expensive calculation of exposure indicators: identifying all parts of a building. This benchmark should also be applied to the bounding box of Berlin, which exhibits very dense mapping with a high level of details. This expensive calculation is currently not part of the processing because it takes more time than the system can afford to keep up with the constant changes in OpenStreetMap.

Some baselines for our benchmarks depend on the existing system hosted by GFZ currently online processing global data and are impossible to measure independently of the live update load; they will be measured once the relevant benchmarks will have been deployed on the project infrastructure. The baselines that are known are the following: the worldwide update cycle is limited to a maximum of one million buildings updated per period of 24 hours on the reference system, and the last benchmark current performance is too low to be active in the update cycle.

4 LINKS WITH OTHER WORKPACKAGES

4.1 ACCELERATION OPPORTUNITIES

A study was made of heterogeneous acceleration opportunities for the pilot, focusing on acceleration of individual components or applications. This study ended with an estimate that the data management and manipulation of the flow would be the best candidate for that acceleration. The main computational load, TsunAWI, being characterised by computation on a unstructured mesh and double floating point calculations, can be considered as already highly optimised to minimise computations, and the irregularity of memory accesses makes acceleration by GPUs or FPGAs a difficult target.

Two parts of the pilot tasks are considered as being able to benefit from heterogeneous acceleration in the pilot:

- Extracting from the applications, mostly the TsunAWI tasks, parts that are oriented towards data manipulation such as output data interpolation, so that they appear as independent tasks that can be accelerated with a GPU or a FPGA, and scheduled on a node that has the proper acceleration resources by the orchestrator.
- Acceleration for the geographical database operations of the flow. That second part would rely on distribution of the data and workload over multiple systems, such as with the GreenPlum[3] system which allows for the consolidation of multiple instances of the GIS system supporting the OpenBuildingMap database, each instance of the GIS (here, PostGIS) being run on a different resource.

These results have been communicated to the LEXIS co-design activity, and summarized in deliverable D2.1 [5].

Moreover, as explained in [1], the implementation and the orchestration of the pilot tasks are a strong component of the overall acceleration targeted for the complete pilot execution, with very significant gains in delay expected at certain points such as in the satellite-based emergency mapping tasks (see [1]).

4.2 DATA MANAGEMENT

Most datasets used by the pilot and benchmarks are public or will be. A detailed description of the datasets is in deliverable D9.2 [3].

[3] https://greenplum.org/

5 CONCLUSION

The presented set of benchmarks and the suggested approaches to increase the processing speed of the earthquake and tsunami pilot will help make the pilot respond to earthquake and tsunami disasters in a timely fashion for the various stakeholders, from resilience managers to rescue workers. Some of the benchmarks can be handled and improved independently while others will need a setup of the processing system for the global building stock because all further optimizations in processing speed need to consider the real amount of data as the system currently does not scale. We expect significant improvements over the LEXIS project duration to all of the critical components.

REFERENCES

[1] T. Goubier, “Earthquake and Tsunami workflow leveraging the modern HPC/Cloud environment in the LEXIS project,” in *proceedings of 22nd International Conference on Network-based Information Systems (NBIS’19)*.

[2] A. Babeyko, A. Hoechner and S. Sobolev, “Modeling tsunami generation for local tsunami early warning in Indonesia,” *Geoph. Res. Abs.,* vol. 10, 2008.

[3] LEXIS, Deliverable, D2.1 Pilots needs / Infrastructure Evaluation Report, 2019.

[4] M. Riyadi, „APPLICATION EARLY WARNING SYSTEM IN INDONESIA,“ *Agency for Meteorology, Climatology and Geophysics of the Republic of Indonesia (BMKG),* 2015.

[5] LEXIS, Deliverable, D9.2 IPR and Data Management Approach, 2019.

[6] V. H. Silva, M. Crowley, D. Pagani and R. Pinho, “Development of the OpenQuake engine, the Global Earthquake Model's open-source software for seismic risk assessment,” *Natural Hazards,* vol. 72, no. 3, pp. 1409-1427, 2014.

A NBIS PAPER

Joined in annex to this report is the publication describing the pilot, which has been accepted in the proceedings of the 22nd International Conference on Network-Based Information Systems (NBiS-2019), to be held September 5-7 September 2019, Oita University, Oita, Japan.

Earthquake and Tsunami workflow leveraging the modern HPC/Cloud environment in the LEXIS project

Thierry Goubier, Andrea Ajmar, Carmine D'Amico, Paul Dubrulle, Susanna Grita, Stephane Louise, Jan Martinovič. Tomáš Martinovič, Natalja Rakowsky, Paolo Savio, Danijel Schorlemmer, Alberto Scionti and Olivier Terzo

Abstract Accurate and rapid earthquake loss assessments and tsunami early warnings are critical in modern society to allow for appropriate and timely emergency response decisions. In the LEXIS project, we seek to enhance the workflow of rapid loss assessments and emergency decision support systems by leveraging an orchestrated heterogeneous environment combining high-performance computing resources and Cloud infrastructure. The workflow consists of three main applications: Firstly, after an earthquake occurs, its shaking distribution (ShakeMap) is computed based on the OpenQuake code. Secondly, if a tsunami may have been triggered by the earthquake, tsunami simulations (first a fast and coarse and later a high-resolution and computationally intensive computation) are performed based on the TsunAWI simulation code that allows for an early warning in potentially affected areas. Finally, based on the previous results, a loss assessment based on a dynamic exposure model using open data such as OpenStreetMap is computed. To consolidate the workflow and ensure respect of the time constraints, we are devel-

Thierry Goubier, Stephane Louise, Paul Dubrulle
CEA, List, PC 172, F-91191 Gif-sur-Yvette Cedex, France, e-mail: firstname.lastname@cea.fr

Jan Martinovič, Tomáš Martinovič
IT4Innovations, VSB - Technical University of Ostrava, 17. listopadu 2172/15, 708 00 Ostrava-Poruba, Czech Republic e-mail: firstname.lastname@vsb.cz

Andrea Ajmar, Susanna Grita
ITHACA, via P. C. Boggio 61, 10138 Torino, Italy e-mail: firstname.lastname@ithaca.polito.it

Natalja Rakowsky
Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Am Handelshafen 12, Bremerhaven Germany e-mail: natalja.rakowsky@awi.de

Alberto Scionti, Olivier Terzo, Carmine D'Amico, Paolo Savio
LINKS Foundation, via P. C. Boggio 61, 10138, Torino, Italy e-mail: firstname.lastname@linksfoundation.com

Danijel Schorlemmer
GFZ German Research Centre for Geosciences, Telegrafenberg, 14473 Potsdam, Germany e-mail: ds@gfz-potsdam.de

oping an extension of a time-constrained dataflow model of computation, layered above and below the workflow management tools of both the high-performance computing resources and the Cloud infrastructure. This model of computation is also used to express tasks in the workflow at the right granularity to benefit from the data management optimisation facilities of the LEXIS project. This paper describes the workflow, the computations associated and the model of computation within the LEXIS platform.

1 Introduction

Simulations in High Performance Computing (HPC) have long been capable (and used) to simulate natural disasters, from earthquakes, tsunamis, floods, forest fires to pest invasions and more. Additionally, geographic information systems have been used to map the extent of natural disasters and compute their effects on people and buildings, infrastructure, and economy, by adding geo-referenced information onto the natural disaster simulation results.

Natural disasters are, however, complex phenomena to model and simulate. Local and small-scale irregularities, such as varying soil and sub-soil conditions or the shape of a bay, can have a huge impact on the shaking of an earthquake or the run-up height or inundation of a tsunami, respectively. To handle that, a regular increase in computational power and precision of the representation of the environment are used, which leads to the main difficulty of this domain: the time needed to simulate is incompatible with the time frame necessary for rapid and effective use of the results.

Multiple pragmatic concepts have been used to compensate for this. The tsunami early warning system for Indonesia, InaTEWS, employs the Tsunami Observation And Simulation Terminal (TOAST)[1] decision support system, connected to both a database of high-quality scenarios precomputed with TsunAWI and to the high-speed online algorithm easyWave with simpler model physics and coarse resolution. This allows for rapid tsunami warning announcements within minutes after an earthquake. While TsunAWI scenarios include inundation, e.g. for planning evacuation routes, they are based on predefined earthquake sources. In contrast, the real-time simulation easyWave starts with an earthquake source determined from the actual measurements and thus better assesses the actual situation, but, as a trade-off, easyWave can only deliver the wave height at the coast without inundation.

Disaster simulation results do not only provide the important early warnings but also valuable information and guidance for short-term emergency response (e.g. personnel to mobilize, shelter to be provided) and long-term disaster recovery actions (e.g. future supplies to order, disruptions to expect, disaster relief funding). Results from simulations have to be available at the right time for those decisions to be

[1] https://www.gempa.de/products/toast/

taken with the best available knowledge, knowing that on-the-ground information will only become available after actions had to be taken to mitigate the disaster.

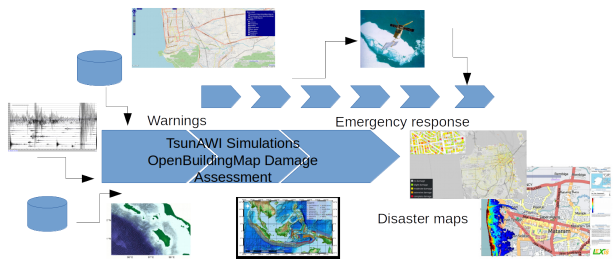

Fig. 1 A high level view of the earthquake and tsunami pilot flow.

To bridge the gap between detailed disaster simulations and rapid emergency decision support and to ensure that emergency response systems (ERS) have the best available information when needed, LEXIS combines knowledge and actors from different domains: natural disaster scientists and code optimization specialists. To achieve this, we plan to combine an earthquake damage assessment model and a tsunami simulation code under a single model of computation, linked to a satellite-based emergency mapping process, as shown in Fig. 1. This system is deployed on the LEXIS platform [1], crossing boundaries between High Performance Computing resources and the Cloud, and linking in data management acceleration technologies. We will detail the objectives, followed by a description of the workflow and its different components.

2 Objectives

In the LEXIS project, we are focused on innovating and accelerating the computation process of the ERS. The objectives are to leverage the LEXIS platform and create a new workflow for emergency response systems based on this platform. The main points of the LEXIS platform are federated HPC and Cloud resources, enhanced by the burst buffer technology.

This is comprised of three pillars:

- Infrastructure,
- Orchestration,
- Data management.

In practice, the LEXIS orchestration tool will distribute the computational resources, based on the available computational resources and data locality, to optimize the computational workflow and provide timely results, taking into account the time constraints expressed in the emergency workflow. Additionally, the orchestrator will know and allocate the parts of the computation that should be sent to the

HPC clusters and which should be sent to the Cloud infrastructure. From the point of the emergency response system, multiple TsunAWI simulations should be computed by the HPC resources, leveraging the burst buffer technology, and the data post-processing should be computed in Cloud, sharing the data from the simulation. Such distribution of the resources should provide the best efficiency in resource usage and data management.

From the point of view of the workflow itself, the objectives are the ability to express time constraints and time-based decisions in the workflow. Time constraints on the production of results, that will drive the orchestration allocation of tasks to resources to ensure the required computations are done on-time. And time-based decisions, allowing the workflow to choose the best results at a point in time, and to manipulate tasks to ensure the most efficient use of the available computing resources given the time available. We expect the following two hypotheses to hold true in the context of this (and similar) workflows:

- Precise simulations will be too costly to prove that they will always end on-time, both for the complexity of those simulations, and for the necessary short time margin to ensure a worst case run ends before the deadlines, so wasting a huge amount of computing resources if the orchestration does not accommodate for simulations that may be unable to provide results before a deadline.
- Compute resources available to the workflow will be dynamic, with a reserved amount possibly guaranteed by contract between an infrastructure provider and civil authorities, and additional resources added on-line during the course of the emergency, mandating an elastic orchestration of tasks.

This is reflected in how the LEXIS technologies will be supporting this workflow, and how the workflow orchestration will be implemented. Knowledge of the compute characteristics of the various tasks of the workflow will also play a significant role, by allowing the manipulation of their parameters to ensure the best fit for the available computation and time resources.

3 The Earthquake and Tsunami Workflow

The earthquake and tsunami workflow is a combination of the OpenQuake code computing the shaking distribution (ShakeMap) of an earthquake, the tsunami simulation code TsunAWI, the dynamic exposure model, the loss assessment code, the satellite-based emergency mapping (SEM) process, and the HeScade time constrained dataflow model of computation, this deployed on the LEXIS platform.

The workflow is both event-triggered and permanently online and requires: provision and update of the best available estimates and results at all time, and provision of the best results so far at specific points of time after the event (those precise points will be defined in the course of the project by interacting with emergency response agencies).

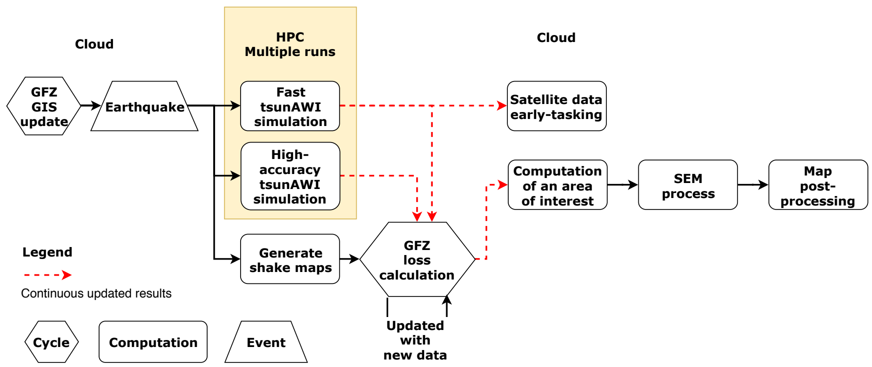

Fig. 2 ERS computational workflow on LEXIS infrastructure

The permanently online component of the workflow is the dynamic exposure model based on OpenStreetMap, which updates itself constantly from the crowd-sourced OpenStreetMap database in real-time, so as to benefit from up-to-date information.

All other parts are triggered upon reception of an earthquake event as shown in Fig. 2. Computations of the shaking distribution and the fast tsunami simulations are started immediately; more precise simulations are launched once more detailed information on the event is available. Fast simulations should provide results for emergency warning systems; more precise results will be provided by the precise simulations (which would require more computation resources and time). The results define areas affected by shaking or inundation and are propagated to the remaining flow.

Upon reception of the earthquake ShakeMap, triggering an extraction of the relevant geographical subset of the world from the exposure model which is subsequently used for computing the loss assessment. This loss assessment will be continuously updated when the tsunami simulations provide results about inundated areas, with an eventual extension of the geographical area affected. Outputs are a geographical representation of the loss assessment and summaries.

The Satellite-based emergency mapping process is triggered by the availability of areas of interest. Such areas are determined by the earthquake event and it's ShakeMap and by additional factors: the amount of tentative damage expected at that location, additional disaster information (extent of inundation or presence of inundation), the importance of settlements at that position. The area of interest is updated through the results of the TsunAWI simulation and the damage/loss assessment; fast simulations allow for a shorter time to determine the early areas of interest and better results allow for continuous refinement of the areas of interest. Damage assessment and inundation results are also integrated into the SEM products.

4 The workflow components

4.1 The HeScade model of computation

We selected a model of computation to allow for taking into account the Real-Time constraints of the ERS. To that end, we define a *Model of Computation* (MoC) based on two types of actors: compute actors (tasks or jobs) and control actors, the latter orchestrate how the former would behave, when they start and when they can or must terminate (depending on the associated constraints within the goals of the application). The model of computation also encompasses the data and communication requirements and interactions between the actors, and can be described as belonging to the class of dataflow models of computation, and will be mapped over the LEXIS platform unified orchestrator, as well as within the platform Cloud and HPC components' task runtimes such as HyperLoom [3] or COMPS [2]. To this end, the LEXIS platform orchestrator will be responsible to install the software components on reserved resources, while the runtime tasks scheduler (e.g., HyperLoom or COMPS) will coordinate the execution of compute and control tasks accordingly.

The MoC defines an application as a K set of compute agents, C as a set of control agents, e as a set of communication edges in $K \times K \cup K \times C$ (*i.e.*, either between compute agents or between a compute agent and a control agent), d as a set of control edges in $C \times K$ *i.e.*, exclusively between a control agent and a compute agent. The edges are connected to so-called data ports and control ports of the agents so that:

- Any compute agent possesses at most one control port,
- No control agent can possess a control port,
- Any port is associated with a number of tokens (either data tokens or control tokens) that correspond to a predefined amount of data transfer which usually is an atomic production (for output port) or intake (for input ports).
- Any compute agent can have an arbitrary number of input or output data ports and each port is associated with a fixed series of numbers in $\mathbb{N}^k$ which defines the number of expected data tokens for each firing of the agent. When the last number of tokens from the series is reached, the series is restarted from the start, hence for any agent, the number of tokens consumed or produced on each port is a constant number throughout the super-period that includes all the agents. This ensures SDF [9] and CSDF [6] equivalence.

Contrary to the usual behaviour of processes in SDF or CSDF, agents in our MoC data input ports are not necessarily blocking when the number of expected data-token is not reached. The occurrence of a control token can "awaken" an actor so that depending on the planned behaviour, it can discard some of the input ports and only use the available data-token to run its processing and produce data-tokens on the output ports. Without control ports, compute agents must wait for all the input ports to reach the number of data-tokens specified by the intake behaviour of the agent, as in the SDF MoC.

The advantage of an SDF/CSDF equivalency is that there exists mathematical proof of when an application described this way is self-consistent and will avoid deadlock while running in finite memory. The extension of control ports allows providing real-time constraints to a system in which time-boundedness (for SDF/CSDF) is not guaranteed.

Therefore, by modelling the application with such a MoC, we can ensure that: first, the system can run without deadlocks and in finite memory provided it is well designed (which is a simple mathematical verification) and second, that the correct use of control agent to enforce latency deadlines to the processes being fired by the orchestrator (see [10] and [4]). With at least one simulation with a parameter set allowing to meet the deadline with the initial resources reserved for the workflow, the availability of the outcome of at least this simulation is guaranteed at the deadline. At the specified deadline, the best available result is identified and can be forwarded to the next step in the workflow.

For the orchestrator engine managing the resources of the platform, this *a priori* knowledge of the availability of results makes the task of allocating resources easier. The resources reserved in case of an alert are used to start the simulation with guaranteed results at the deadline. Unused reserved resources can be used to start additional simulations. Also, considering that additional resources can become available after an alert, it is interesting to introduce a work-stealing approach that is driven by the data flow dependencies, as in [7]. In any case, the communication resources for the selected result can be reserved independently.

4.2 TsunAWI

TsunAWI was developed in the framework of the German-Indonesian Tsunami Early Warning System (GITEWS, 2005-2011, funded by the German Federal Ministry of Education and Research, [12]) and simulates all stages of a tsunami from the origin, the propagation in the ocean to the arrival at the coast and the inundation on land [11]. It solves the non-linear shallow water equations (SWE) on unstructured triangular finite elements that allow covering the coastal areas with a high resolution, while the long tsunami waves in the deep ocean can be represented by a coarse mesh, thus saving computation time and memory. The quality of the triangulation is crucial for the model results. We usually employ a mesh generator based on Triangle [14] with additional smoothing steps as described in [5]. Starting from a model domain defined within a topography/bathymetry data set, the mesh generator builds a mesh based on refinement rules depending on the water depth, the gradient of the bathymetry, and user defined criteria like minimum and maximum resolution and optional foci on regions of interest.

The scenarios for the Indonesia tsunami early warning system InaTEWS have a resolution of 12km in the deep ocean, 150m at the coast, and 50m in project regions and around tide gauges. In hindcasts of real events, the resolution of 50m is sufficient to simulate realistic waveforms at most tide gauge locations. For inundation studies, however, the horizontal resolution should be as fine as 20m, as shown in [8]. Accordingly, the setup for LEXIS consists of a regional mesh with 20m resolution in the city of Padang, 150m elsewhere at the coast, and 5km in the ocean, see figure 3.

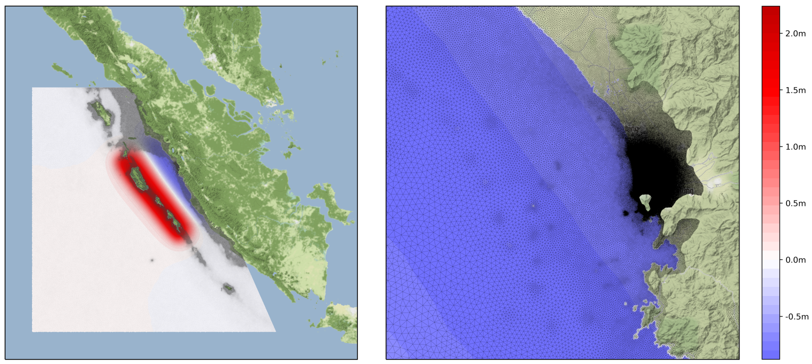

Fig. 3 Mesh with an initial condition for the simulation of tsunami inundation in the city of Padang. Map tiles by Stamen Design, under CC BY 3.0. Data by OpenStreetMap, under ODbL

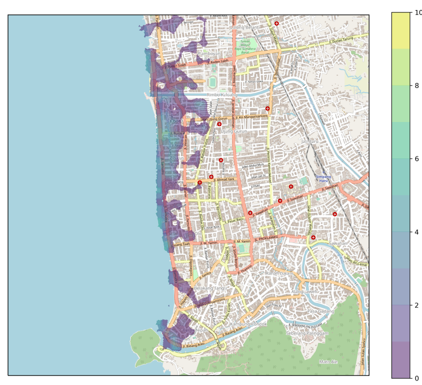

Fig. 4 Inundation simulation for Padang Map tiles by OpenStreetMap

4.3 Global Dynamic Exposure

Detailed understanding of local risk factors regarding natural catastrophes requires an in-depth characterization of the local exposure. In LEXIS we aim at using an exposure model [13] providing exposure and vulnerability indicators for all building footprint present in OpenStreetMap (OSM). OSM is the rich and constantly growing geographical database, that contains more than 5 billion geographical nodes, more than 1/3 of a billion building footprints (growing by more than 100,000 per day), and a plethora of information about school, hospital, and other critical facilities.

To ensure timely delivery of exposure and vulnerability indicators for an area affected by an earthquake or a tsunami, all indicators need to be precomputed and kept up-to-date. This requires a computation of indicators anytime a building is added, changed, or deleted from OSM or a change is affecting the indicators of a building. This near-realtime processing is computationally demanding because in the order of 1 million buildings globally need to be (re-)assessed daily. With this approach, we increase the resolution of existing exposure models from aggregated exposure information to building-by-building vulnerability.

The impact of the LEXIS project on the exposure model and loss assessment based on the exposure model will be twofold:

- The process is data and compute intensive, and the project platform will be used to accelerate and allow the process to scale. Key elements are the need to keep the database up-to-date all the time so as not to have to base a damage estimate on outdated data, and the ability to extract the impacted area from the global database without delay upon reception of an event.
- In emergencies and for disaster response, having the possibility to ensure that results are delivered at the point where they are needed, and so benefit from acceleration and orchestration technologies ensuring urgent tasks are deployed timely and on the right resources

4.4 Integration into an Satellite-based Emergency Mapping workflow

The ultimate goal of an SEM mechanism is to improve disaster relief effectiveness and thus to help reduce suffering and fatalities before, during, and after a disaster event occurs [15]. Specifically dealing with the response phase immediately after a disaster, which typically lasts from several days to a few weeks, time is the major constraint: this is where automatic models and procedures potentially provide the greatest benefits. The Copernicus Emergency Management Service (CEMS), the European Commission SEM mechanism, is composed of an on-demand mapping component providing rapid maps for emergency response and risk & recovery maps for prevention and planning. Two are the specific CEMS phases that can be significantly improved by the exploitation of models and algorithms integrated into the LEXIS platform: a) early-tasking of satellite acquisitions, exploiting the output of the TsunAWI model and b) automatically generate a first estimate product (FEP), combining TsunAWI inundation outputs with building exposure data.

Most of earth observation satellite platform acquire images on request and have specific cut-off time to program an acquisition: missing this cut-off time would lead to a missing opportunity for data acquisition and to at least a 24 hours delay. TsunAWI estimated wave height on the coast, crossed with some globally available exposure datasets, such as population places or population distribution, can be used to define and prioritize a set of Areas of Interest (AoI) polygons, to be submitted to the satellite data provider in the shortest time frame possible.

A First Estimate Product (FEP) provides a very fast yet rough assessment of the most affected locations. Coastal area inundation extent and maximum water height can be integrated with available exposure data at a single building level, in order to quickly generate first damage estimates, while waiting for the results of the final analysis which is more time consuming. In big disasters, like tsunamis, FEP allows understanding the situation quickly and supports identifying focus areas for further image tasking and analysis.

Fig. 6 displays the benefit of the LEXIS workflow, compared with a standard one triggered by a user activation request: in the reproduced scenario, thanks to the early-tasking, the first post-event image is expected to be delivered 1 day in advance. Furthermore, the FEP product is generated 1.5 days earlier, thanks to the availability of the TsunAWI inundation and the building exposure data.

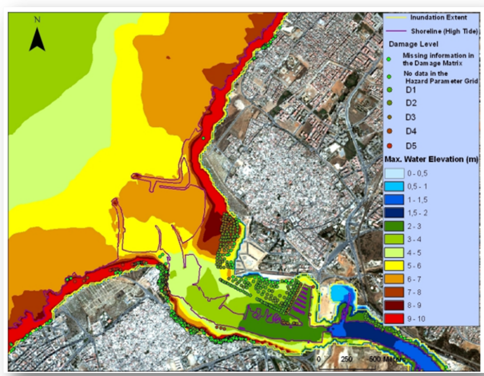

Fig. 5 Example of a FEP based on inundation extent and maximum water height combined with building exposure

5 Conclusion

We have shown how, in the LEXIS project, an earthquake and tsunami workflow will be built, leveraging an orchestrated heterogeneous environment combining high-performance-computing and Cloud resources.

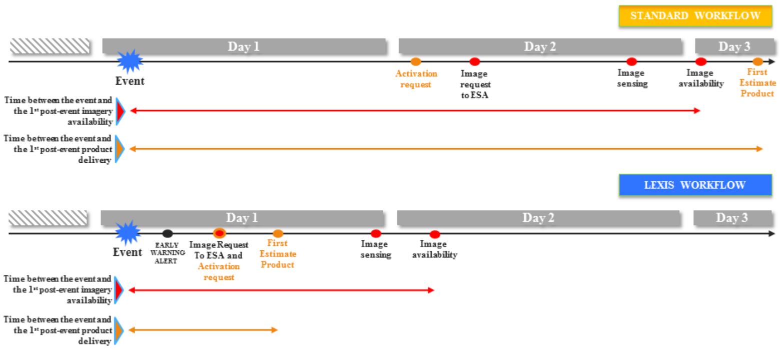

Fig. 6 Comparison, in terms of product timeliness, of a standard workflow and one based on LEXIS workflow

The tsunami simulation code TsunAWI will be used to provide both fast on-line simulations and accurate compute intensive simulations after an earthquake event. The dynamic exposure model, kept up-to-date with the latest information from the OpenStreetMap database, will be used to compute damage and loss assessments based on the earthquake ShakeMap and the tsunami inundations. And a satellite-based emergency mapping process will produce early, accurate maps of the affected areas. A time constrained dataflow model of computation will be used to ensure the orchestration of the workflow will respect the time constraints and optimize for the computing resources available.

This will be deployed over the three key technologies of the LEXIS project, that are the infrastructure, the orchestration and the data management. Infrastructure services will be ensured by the LRZ and IT4Innovations centres, and heterogeneous acceleration capabilities will be employed to accelerate key points of the workflow.

Acknowledgements This work was supported by the LEXIS project - the European Unions Horizon 2020 research and innovation programme under grant agreement No. 825532.

References

1. Olivier Terzo Etienne Walter Marc Levrier Stephan Hachinger Donato Magarielli Thierry Goubier Stephane Louise Antonio Parodi Sean Murphy Carmine DAmico Simone Ciccia Emanuele Danovaro Martina Lagasio Frederic Donnat Martin Golasowski Tiago Quintino James Hawkes Tomas Martinovic Lubomir Riha Katerina Slaninova Stefano Serra Roberto Peveri Alberto Scionti, Jan Martinovic. Hpc, cloud and big-data convergent architectures: The lexis approach. In *Conference on Complex, Intelligent, and Software Intensive Systems (CISIS)*, (forthcoming) 2019.

2. Rosa M Badia, Javier Conejero, Carlos Diaz, Jorge Ejarque, Daniele Lezzi, Francesc Lordan, Cristian Ramon-Cortes, and Raul Sirvent. Comp superscalar, an interoperable programming framework. *SoftwareX*, 3:32–36, 2015.
3. Vojtěch Cima, Stanislav Böhm, Jan Martinovič, Jiří Dvorský, Kateřina Janurová, Tom Vander Aa, Thomas J. Ashby, and Vladimir Chupakhin. Hyperloom: A platform for defining and executing scientific pipelines in distributed environments. In *Proceedings of the 9th Workshop and 7th Workshop on Parallel Programming and RunTime Management Techniques for Manycore Architectures and Design Tools and Architectures for Multicore Embedded Computing Platforms*, PARMA-DITAM '18, pages 1–6, New York, NY, USA, 2018. ACM.
4. XuanKhanh Do, Stéphane Louise, and Albert Cohen. Transaction parameterized dataflow: A model for context-dependent streaming applications. In *2016 Design, Automation & Test in Europe Conference & Exhibition, DATE 2016, Dresden, Germany, March 14-18, 2016*, pages 960–965, 2016.
5. William H. Frey and David A. Field. Mesh relaxation: a new technique for improving triangulations. *Int. J. Numer. Methods Eng.*, 31(6):1121–1133, 1991.
6. R. Lauwereins G. Bilsen, M. Engels and J. A. Peperstraete. Cyclo-static data flow. *IEEE Transactions on Signal Processing*, 44(2):397–408, 1996.
7. Thierry Gautier, Xavier Besseron, and Laurent Pigeon. Kaapi: A thread scheduling runtime system for data flow computations on cluster of multi-processors. In *Proceedings of the 2007 International Workshop on Parallel Symbolic Computation*, PASCO '07, pages 15–23, New York, NY, USA, 2007. ACM.
8. J. Griffin, H. Latief, W. Kongko, S. Harig, N. Horspool, R. Hanung, A. Rojali, N. Maher, A. Fuchs, J. Hossen, S. Upi, S. E. Dewanto, N. Rakowsky, and P. Cummins. An evaluation of onshore digital elevation models for modeling tsunami inundation zones. *Frontieres in Earth Science*, 3(32), 2015.
9. E.A. Lee and D.G. Messerschmitt. Synchronous data flow. *Proceedings of the IEEE*, 75(9):1235 – 1245, sept. 1987.
10. S. Louise, P. Dubrulle, and T. Goubier. A model of computation for real-time applications on embedded manycores. In *Embedded Multicore/Manycore SoCs (MCSoc), 2014 IEEE 8th International Symposium on*, pages 333–340, Sept 2014.
11. Natalja Rakowsky, Alexey Androsov, Annika Fuchs, Sven Harig, Antonia Immerz, Sergey Danilov, Wolfgang Hiller, and Jens Schr"oter. Operational tsunami modelling with TsunAWI recent developments and applications. *Nat. Hazards Earth Syst. Sci.*, 13:1629–1642, 2013.
12. A. Rudloff, J. Lauterjung, and Münch U., editors. *The GITEWS Project (German-Indonesian Tsunami Early Warning System)*. NHESS - Special Issues. Copernicus Publications, Göttingen, 2009.
13. D. Schorlemmer, T. Beutin, N. Hirata, M. Wyss, F. Cotton, and K. Prehn. Global Dynamic Exposure and the OpenBuildingMap - Communicating Risk and Involving Communities. In *EGU General Assembly Conference Abstracts*, volume 20 of *EGU General Assembly Conference Abstracts*, page 12871, April 2018.
14. Jonathan Richard Shewchuk. Triangle: Engineering a 2D Quality Mesh Generator and Delaunay Triangulator. In Ming C. Lin and Dinesh Manocha, editors, *Applied Computational Geometry: Towards Geometric Engineering*, volume 1148 of *Lecture Notes in Computer Science*, pages 203–222. Springer-Verlag, May 1996. From the First ACM Workshop on Applied Computational Geometry.
15. Stefan Voigt, Fabio Giulio-Tonolo, Josh Lyons, Jan Kučera, Brenda Jones, Tobias Schneiderhan, Gabriel Platzeck, Kazuya Kaku, Manzul Kumar Hazarika, Lorant Czaran, Suju Li, Wendi Pedersen, Godstime Kadiri James, Catherine Proy, Denis Macharia Muthike, Jerome Bequignon, and Debarati Guha-Sapir. Global trends in satellite-based emergency mapping. *Science*, 353(6296):247–252, 2016.