D6.7 – Third annual report on training, standardisation, collaboration, dissemination, and communication

Grant Agreement	676547
Project Acronym	CoeGSS
Project Title	Centre of Excellence for Global Systems Science
Topic	EINFRA-5-2015
Project website	http://www.coegss-project.eu
Start Date of project	October 1, 2015
Duration	36 months
Deliverable due date	30.09.2018
Actual date of submission	30.09.2018
Dissemination level	Public
Nature	Report
Version	2 (after internal review)
Work Package	6
Lead beneficiary	GCF
Responsible scientist/administrator	Sarah Wolf (GCF)
Contributors	Sebastian Böhne (UP), Michael Gienger (HLRS), Fabio Saracco (IMT), Jette von Postel (GCF)

Internal reviewers	Alexandra Gens (DIALOGIK), Bastian Koller (HLRS)
Keywords	Training, Standardisation, Event Management, Collaboration, Dissemination, Communication
Total number of pages:	31

The CoeGSS ("Centre of Excellence for Global Systems Science") project is funded by the European Union. For more information on the project please see the website http:// http://coegss-project.eu/

Version History

	Name	Partner	Date
From	Sarah Wolf	GCF	
First Version	for internal review		09/2018
Second Version	for internal review		9/2018
Third Version	for submission		9/2018
Reviewed by	Alexandra Gens Bastian Koller	DIA HLRS	9/2018 9/2018
Approved by	ECM	ATOS,UP,USTUTT	9/2018

Abstract

This meta-deliverable presents the state of work in several tasks of WP6 "Awareness Creation and Community Support" at the end of the third project year; in particular, it comprises the activities and achievements of the tasks

T6.2 Event Management and Collaboration

T6.3 Training

T6.4 Standardisation

T6.5 Dissemination and Communication.

These achievements comprise events organised and collaborations carried on over the third project year, the formulation of a curriculum for HPC experts tailored to GSS needs and the identification of best practices in training courses as well as training material produced, monitoring of standards and benchmarking activities, and a report on dissemination of all elements of CoeGSS.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ABM	Agent Based Model
CART	Classification and Regression Trees
CKAN	Comprehensive Knowledge Archive Network
CNR	National Research Council (Consiglio Nazionale delle Ricerche), Italy
CoeGSS	Centre of Excellence for Global Systems Science
CPGC	Computing Power for Global Challenges
DSL	Domain Specific Language
EC	European Commission
EDGE	Exascale, Data and Global Evolutions
GCF	Global Climate Forum e.V.
GG	Green Growth
GS	Global System
GSS	Global Systems Science
GSSE	GSS-expert
HLRS	Höchstleistungsrechenzentrum Stuttgart
HPC	High-Performance Computing
HPCE	HPC-expert
HPDA	High-Performance Data Analytics
HWRF	Hurricane Weather Research and Forcasting
IPF	Iterative Proportional Fitting
MOOC	Massive Open Online Course
MoTMo	Mobility Transition Model
ODD	Overview, Design concepts, Detail
PSNC	Poznan Supercomputing & Networking Center
SIS	Synthetic Information Sytem
SP	Synthetic Population
TDP	Thermal Design Power
WP	Work package

1 Introduction

This deliverable presents the third project year progress of tasks T6.2-T6.5 in WP6 "Awareness Creation and Community Support" of the Centre of Excellence for Global Systems Science (CoeGSS):

T6.2 Event Management and Collaboration

T6.3 Training

T6.4 Standardisation

T6.5 Dissemination

Its companion deliverable, D6.4, reports on T6.1, community building. As planned, these tasks continued the dissemination of the CoeGSS approach and its results, and interacted with external stakeholders through events and training activities. Section 2 reports on events organised by CoeGSS, such as the “International Conference: Computing Power for Global Challenges”. Section 3 presents the curriculum that the requirements analyses in the previous iterations of this delibverable worked towards, and reports on training activities and materials produced. Section 4 concludes the task on standardization, and Section 5 assembles the dissemination report of the third and last project year, before Section 6 concludes.

2 Event Management and Collaboration

2.1 Events organized

2.1.1 CPGC Conference

The "International Conference: Computing Power for Global Challenges" was held at the IMT School for Advanced Studies Lucca, Italy, on the 24th and 25th of October 2017. The conference was organized by CoeGSS in collaboration with the Dolfins project (https://simpolproject.eu/); the website (developed by Top-IX) can be found at http://cpgc.coegss.eu/homepage.

The Top-IX crew recorded interviews of some of the speakers and was responsible for the online live recording. The audience was composed of around 40 persons, including the most reknown experts in different GSS fields and from HPC.

On the first day, international experts took part in discussions in dedicated panels on four selected global challenges: developing a sustainable and resilient global financial system, addressing the daunting risks of pandemics, transforming the fossil-fuel based global mobility system, and creating forms of democracy adequate to the age of the digitalization. On the second day, the debate on these topics was furthered in roundtables. The aim of the conference was to trigger the discussion among experts coming from different fields, showing the last achievements and presenting open challenges.

On the 24th, Luis Neves (Global e-Sustainability Initiative), Carlo Jaeger (Global Climate Forum and University of Potsdam) and Michael Resch (High Performance Computing Centre Stuttgart) introduced the event, after a greeting from Guido Caldarelli (IMT School for Advanced Studies, Lucca) and the director of IMT, Pietro Pietrini. Robert Axtell (George Mason University), Sanzio Bassini (Super Computing Applications and Innovation) and Carlo Jaeger discussed about the role of HPC for Global Challenges in the first panel session.

In the following topical sessions, a discussion about the use of supercomputers for pandemic studies was held between Alessandro Vespignani (Northeastern University, Boston) (remotely), Chris Barret (Virginia University), Madhav Marathe (Virginia University) and Bruno Gonçalves (New York University). Carlo Jaeger, Stefano Battiston (University of Zürich) and Marco Bardoscia (Bank of England) discussed about the improvements obtained by using network approaches for the study of the resilience of financial systems and how they are implemented in national banks.

Sonia Yeh (Chalmers), Joshua Auld (Argonne National Laboratory) and Sarah Wolf (Global Climate Forum) presented results of implementing models using HPC for the study of strategies for a sustainable mobility transition. In the last panel, Guido Caldarelli, Erika Widegren (ReImagine Europa), Raffaele Perego (Institute of Information Science and Technologies, CNR) and Miguel Maduro (European University Institute) addressed the problem of an adequate democracy in a digitalised environment and the development of infrastructures and institutions addressing these issues.

The second day was devoted to round tables, organised in parallel sessions. The contributions to the conference can be found on the program wepage (http://cpgc.coegss.eu/program), while the dedicated interviews are in the speakers page (http://cpgc.coegss.eu/speakers).

Overall, the conference provided an opportunity of community building, dissemination, collaboration, and initiating new collaborations that only face-to-face events can present.

2.1.2 Statistical Physics for Financial & Economics Networks: NetSci2018 satellite

One of the potential future applications of GSS on HPC, identified in the Future Applications Task in WP4, concerns financial stability. There, financial and economic networks play an important role. To establishing contact with scholars and instititutions on this topic, on the 12th of June 2018, the NetSci2018 satellite "Statistical Physics for Financial & Economics Networks" was held. The aim of the satellite was to collect the latest achievements about the application of complex network analysis on financial and economics networks. The audience was composed of about 50 persons, being mostly networks study experts, both from national banks (as the Bank of England) or reknown research institutions. The satellite was organized by IMT School for Advanced Studies Lucca and its website can be found at https://sites.google.com/imtlucca.it/spfen3-netsci2018/home. Marco Bardoscia (Bank of England), Monica Billio (Università Ca' Foscari), Damien Challet (CentraleSupélec) and Andrea Gabrielli (Institute for Complex Systems – CNR) were invited as speakers. Multiplex approaches for the study of financial and economic networks were discussed in the first sessions. In the third one, entropy approaches to network reconstruction were presented. Finally, several analyses of the dynamics of networks were proposed.

2.1.3 Stakeholder Workshop on Mobility in Turin

On the 27th of June 2018, TOP-IX and GCF organized a short workshop titled "What if… – Thinking a sustainable mobility transition" at the premises of Talent Garden, a coworking space that aims to create local, vibrant, globally connected campuses that empower digital & tech communities (https://talentgarden.org/en-it/). The workshop introduced the CoeGSS approach and tools (in Italian), taking as an example the Green Growth pilot's synthetic information system, in order to discuss how this kind of work can be of use for decision makers in the policy and business worlds. After Leonardo Camiciotti briefly introduced CoeGSS, Sarah Wolf gave a presentation on MoTMo, from the ideas behind it to the possibilities of interactively visualizing model outputs in decision support discussions. The audience was small (about 10 people) but very interested. Since the feedback gathered is more important to the work in WP2, it is described in the related deliverable D2.4.

2.2 Collaboration

During the last project year, partners continued collaboration with other groups and initiated new contacts. In particular, ATOS engaged in discussions with the FORTISSIMO II (https://www.fortissimo-project.eu/about/fortissimo-2) project as well as with MSO4SC (http://mso4sc.eu/), exploring similarities and differences with respect to the portal development of each of the projects, so that CoeGSS could learn from the experience made in these other projects and vice versa. Chalmers benefited from synergies with other projects that some of the CoeGSS researchers were involved in at the same time: the GRACeFUL project (https://www.graceful-project.eu/) on climate resilient urban design, as well as a project on electric vehicles in Gothenburg, that uses synthetic populations and agent-based modelling, and one on energy consumption in buildings; the latter two imply collaboration with other Chalmers departments and Virginia Tech. UP initiated new contacts with other participants at a workshop on "Simulations on high-performance computers", organised by the Institute for Advanced Sustainability Studies and the Berlin-Brandenburg Academy of Sciences and Humanities, and discussed code improvements in the IFIP Working Group 2.1 on Algorithmic Languages and Calculi (ifipwg21.org/). HLRS created more CoeGSS visibility, especially in Germany and the state of Baden-Wuerttemberg, by knowledge transfer activities with bwHPC-S5 (http://www.bwhpc-c5.de/). GCF discussed possible mathematical analysis of the Green Growth pilot model with Christof Schütte (ZIB) and Stefan Klus (Free University Berlin). For the analysis of Financial Data, IMT is continuing its collaboration with Stefano Battiston (University of Zürich), Laura Silvestri (Bank of England), Paolo Barucca (LIMS, London) and Kartik Anand (Deutsche Bundesbank).

3 Training

3.1 GSS-Curriculum (for HPC-experts)

As CoeGSS brought together two previously unrelated communities, an initial role of the training task was that of introducing GSS to HPC-experts and vice versa; this was discussed in D6.5 at the end of the first project year. Then, a requirement analysis of competencies related to the CoeGSS workflow was presented in D6.6, at the end of the second project year. As the CoeGSS workflow combines GSS and HPC, the competencies identified also did. This final deliverable now presents the "curriculum for future HPC technology leaders related to the GSS needs" that the CoeGSS training task promised to deliver, and hence focuses on presenting a GSS-curriculum. While HPC-skills a GSS-expert (GSSE) should acquire and HPC-tools he has to know are equally relevant, these are discussed (explicitly and implicitly) in various other deliverables, e.g., in relation to the pilot studies. Also, the availability of courses on HPC (that interested GSS experts can attend) is enourmous, while for learning GSS, very few courses exist. The contribution of CoeGSS therefore focuses on the latter.

The curriculum developed here – at a conceptual stage – has two parts. The first of them is designed for all persons interested in GSS while the second specifically adresses HPC-experts (HPCEs). The former part comprises a general introduction to GSS (3.1.1) and the treatment of two fundamental aspects of GSS, namely policies and policy makers (3.1.2) as well as Synthetic Information Systems (3.1.3). The HPC-specific part shows the HPCEs in which domains of GSS their expertise will be particularly helpful (3.1.4). Furthermore – and not less important – the HPCEs have to be aware that the GSS work process differs from that in more traditional HPC fields (3.1.5).

The reader should note that this division into two parts is for didactical purposes only. It does not represent a suitable division of the GSS-subject. In particular, there are no clear borders between HPC-related and HPC-unrelated aspects in GSS.

In the following, many aspects are listed the person interested in GSS must know or has to be aware of. For the sake of readability phrases like "the person interested in GSS has to know that x" or "the HPCEs have to be aware of x" will be omitted. Instead, only "x" will be stated. Other more active skills, by contrast, are always explicitly mentioned.

3.1.1 Introduction to GSS

An important task is to raise the awareness for the reasons GSS is needed. First, this includes an understanding that problems related to global systems (GSs) – used up to then as intuitive concept – cannot be treated in the same way as localised ones, like dealing with car accidents or fire, and cannot yet be managed as well as these simpler risks. A second particularity is that failures of global systems are always big failures. This leads to an exceptional responsibility in the investigation of global systems. Then, GSS is needed because traditional approaches, which tried to improve the different components of global systems in isolation, have failed.

Another problem with traditional approaches is that they cannot cope with the objects of global systems, namely the agents, who usually react on insights won, the actions announced etc. Note furthermore that important aspects of many of the problems related to global systems are based on the prisoners' dilemma. As a consequence, GSS is important for addressing the question of the right investments to reach new equilibria.

Next, those interested in GSS should be aware of the context in which it is used. Often policy makers need to come to a decision in situations in which global systems are involved. The GSS scientist is needed to support this decision process by pointing out different policy options and their (possible) outcomes in an easily digestible form. The GSS-scientists in turn usually need help from HPCEs to conduct big simulations, particularly requiring the management of a huge amount of data. Together the GSSEs' and HPCEs' work needs to enable the policy maker to assess whether a particular policy will have the desired outcomes and no (or at least only bearable) harmful side effects. It should be noted that neither GSSEs nor HPCEs decide which outcomes are the desirable ones, since this is the job of the policy maker only. However, sometimes policy-makers are not far-sighted enough. Therefore, it is one of the tasks of the GSS-scientists to create awareness for GS-problems. The HPCEs have to know that in such cases possible policies should be investigated in advance although no policy-maker explicitly asked for this.

Another important aspect of GSS is the object of study: global systems. First, systems can be conceived as organised assemblies, i.e. as a conglomerate of objects with connections between them. Furthermore, global systems operate on different levels. In addition to that they comprise a great many connections and this usually leads to situations in which everything affects everything. In particular, little changes in the conditions of the system might lead to very different outcomes and investigations about global systems sometimes cannot be restricted to smaller units like countries.

However, one cannot or does not want to operate directly on global systems. This is why Synthetic Information Systems are investigated instead; where the latter can be understood as synthetic, statistical copies of the global systems. In subsection 3.1.3 they will be discussed at length. From a bird's eye perspective, what one should know, however, is just that Synthetic Information Systems together with suitable hardware – which may or may not be HPC machines depending on the global system under consideration and the stage of the investigation –, allow to conduct simulations that are in need of many ressources. Results of the simulation can then be visualised by suitable visualisation-software – which again may or may not be HPC-specific (like COVISE for instance).

3.1.2 Policies and policy makers

Policies can be conceived as designing the future. This design-process is based on beliefs about how the world is, how it ought to be, and how it can be made as it ought to be. The objective quality of a policy depends only on the question whether the intended outcomes and no unintended, harmful outcomes are caused.

The complexity of global systems, in which everything is connected to everything, cannot be conciliated with policies that are based on unreflected data, guesswork, or instinct only. One reason is that people tend to behave worse than the idealistic picture one has of them; this pertains to cognitive and ethical aspects as well. Therefore, the respective policies must be adapted to the real behaviour of humans instead. Another problem related to the investigations of global systems is that they do not result in point-predictions (for example that a particular human will be sick in four weeks). Instead, their outcome is of a statistical nature. Next, it is hard to determine the impact many factors in a global system have. Finally, unforeseen events may change the complete outcome of a global system. To analyse, under these conditions, which policies will be suited and which will not one needs objectivity of the data (with the behaviour of the people), frequent tests to see how stable a global system is, and the ability to draw suitable consequences from statistical results.

Not only the policies are of interest but also people or instances that are able to implement such policies. Of course, politicians are the obvious first example. The media, congregations etc. which are not elected, are another kind of policy makers. Easy to forget but very important is a last group of policy makers: the citizens. Excluding the citizens from the process of opinion making in politics is dangerous since the citizens might then look for those politicians who provide the easiest answers, which then are the only ones they are able to understand. This shows that policies cannot act on people but have to act with people.

One important aspect with respect to policy development is to understand the role feedback plays. On the one hand, it is possible to interfere with the help of negative feedback in a system that shows unwanted behaviour. Of course, it has to be monitored whether the intervention has indeed the desired consequences. On the other hand, there are positive feedback loops, too, i.e. self increasing effects, which can but do not have to be positive (like panic buying). Furthermore, there are situations, like birth and death for instance, with natural positive and negative feedback loops that can balance each other.

3.1.3 Synthetic Information Systems

Synthetic Information Systems (SISs) are the desired tool for persuasive decision making. They provide counter-factual analysis for what-ifs and allow scientists to ''conduct'' experiments, which would be unethical in the real-world. The reason for this capability is their ability to simulate the actions and interactions of autonomous agents. In particular SISs can be used to identify equilibria and emergent patterns (if there is not only an unintelligible mangle). Furthermore, SISs can help to identify those time slots, in which interventions have great consequences. All these achievements can be seen best when one is confronted with actual simulations showing them.

A SIS consists of the following components:

- a conceptual model that explains the (static and dynamic) patterns for the agents,
- data that must be collected and preprocessed,

- a synthetic population (SP) consisting of agents, which in turn are created by instantiating the static patterns with the preprocessed data,
- an agent-based model (ABM) that on the basis of the available data specifies the behaviour and interactions of agents (the dynamic patterns),
- ABM simulation runs that depict potential evolutions of the global system, together with analysis and visualisation of these.

It can be used for the assessment of possible consequences of decisions.

Next, some non-technical details of the respective components should be noted (the technical details will be discussed in subsection 3.1.4.). In the case of the data there are structured resources like census-data and unstructured ones like those that can be found in the web. Those resources may be also about people's activities and interactions. An important task one should be aware of is to ascertain the quality and the relevance of the data. Some of the data is universal in character. It is worth organising such data in a shared pool. Finally, some preprocessing is needed so that the data can be used later on to create synthetic populations.

Synthetic Populations are sets of synthetic agents reproducing statistical distributions of the real-world population for the relevant features. The properties of the agents (like social, health, cognitive, or cultural attributes) are given by the conceptual model. Environment and interaction networks as well as activities can belong to the properties as well. A synthetic network is a virtual and statistical representation of real world contacts and/or social networks, possibly also in the internet. Together, this all forms agent-patterns that are instantiated to agents with the help of the collected and preprocessed data. Note that the virtual agents do not correspond to specific real individuals but are virtual creations that, as a population imitate the real population (or other entities like places, households etc.). As in the case of data the suitability of the conceptual model and the methods to create SPs out of it together with the data has to be tested.

ABMs are developed for investigating systems with agents. ABMs provide simple rules which determine (stochastically and) stepwise the actual behaviour of the agents in the simulations. The rules may represent the learning of an agent or adaptive processes. In addition to that they may describe interactions between the agents and the environment, in which those interactions take place. There can be different ABMs for the same SP and the quality of the respective ABMs has to be tested.

Simulations are a virtual laboratory for investigating possible system evolutions to find out what the strengths and weaknesses of respective policies are. For the simulations to be useful, there must be some visualisation tool, showing ideally not only the results at the end but the intermediate processes as well. There should be an interface that allows the user of such a tool to put a request on the data he or she is interested in. Note that a single simulation is not a trustworthy result. Instead the simulation has to be repeated under the same and with slightly varying starting conditions to see whether the outcome is stable; regardless of possible (minor) errors in the data. Some technique(s) should be used to make the degree of uncertainty more precise (see subsection 3.1.4). It is very important for large scale simulations of global systems to be resource saving. Therefore low-scale simulations have to be done first to determine which parameters are really important and which ones seem to have (hardly) any impact.

3.1.4 HPC-expertise in GSS

When investigating GSs there are a lot of tasks involved for which HPC-expertise is needed. Data collectors and pre-processors must be built and the data must be stored in a suitable way. Conceptual models have to be designed and instantiated via suitable methods and tools with the concrete data at hand to become synthetic populations. In addition to that and again with the help of the data ABMs have to be designed and implemented. Suitable tools for the simulation process are needed; including a (semi)-automatic analysis of the resulting data. Suitable visualisation tools and an interface for users' requests must be available. Therefore, in the following the requirements of the respective tasks will be discussed in more detail.

Data in global systems is usually classified as big, i.e. it usually has a high volume (more than one can store or process in one place), high velocity (real time information) and high variety. So, dealing with this kind of data turns out to be particularly difficult. In particular, investigations are needed to find simple (and cheap) ways to gather good and relevant data. The hope is to develop (semi)-automatic tools for that task. To avoid that the same work is done several times there should be a common pool of such pre-processed data that is needed in the analysis of many GSs, where the pre-processing itself should also be (semi)-automatic (in the CoeGSS portal, CKAN plays this role). Standard formats and meta-data explaining the collected data are required for this. Further problems pertain to bad or incomplete data. As an example of the former, one might need to compare data of the same kind but in different countries in which the data-quality differs considerably. Or there is different data available for the same request. With respect to the case of incomplete data, one problem is that not all data is visible. For instance, one person may have a contagious disease without the corresponding symptoms. Another problem is that data can quickly become outdated. Extrapolation techniques to create new data might help here. If there is no data available for one country (or some other instance) sometimes it might be reasonable to copy the data from a comparable country, in which data is available.

Domain Specific Languages (DSLs) are an important multi-purpose tool for specifications. They can describe constraints, can structure input and output data, facilitate specifications, correctness proofs, and property based testing. A decisive advantage of DSLs is that domain experts using them need not be programming experts. On the technical side, functional languages like LISP, Haskell, Agda, or Idris can be used to compile such DSLs.

HPC-expertise is required when SPs are instantiated. There, algorithms like Iterative Proportional Fitting (IPF) or Combinatorial Optimisation have to be implemented in a way that is suited to be conducted with big data. The same is necessary with respect to the activity and location assignment, which can be realised for instance with CART-based resampling or the fitted value method in the former case and with the gravity model, the trip chaining model, or the radiation model in the latter case.

Dealing with ABMs requires specification and implementation of the model. It must be documented and tested at various stages of its development and the code must then be optimised.

As already mentioned, simulations should not only provide a single result but should be repeated to obtain meta-results about the stability and uncertainty. With respect to the former sensitivity analyses (which parameters are relevant at all) and calibrations (what is the ratio between differences in the input and the ones of the output) must be conducted. Uncertainty will be discussed below separately. Another feature of simulations is that they expose unknown unknowns. These have to be recognised with the help of the computer. Furthermore, simulations can help to identify non-programmed behaviour like congestion behaviour of vehicles on motorways. A technique which is needed for big scale simulations is parallelisation. Here the HPC-expertise is particularly helpful.

It was already discussed that suitable visualisation tools and interaction tools have to be developed. The software must be able to work with all relevant file formats. Remote access for big data simulations must be enabled. Multi-dimensional data must be also visualised in a suitable manner. Best would be an augmented reality realisation.

One aspect that requires HPC-expertise, but that is not restricted to any of the aspects above, is the handling of uncertainty. There are many reasons for faults: inferior data, logical misconceptions in the software, physical errors provoked by the gigantic number of calculations involved etc. DSLs as discussed above are one way trying to avoid faults of the second kind. By contrast, interval arithmetic tries to tame faults of the first kind. Another approach to deal with inferior data is to work with probability distributions and probabilistic algorithms. However, not all faults can be avoided. Instead a reasonable way to deal with such errors is needed. One idea, for instance, is to save (parts of) the result as a snapshot such that a restart – if necessary – can follow from this snapshot (treated under reliability in T3.1 of CoeGSS).

3.1.5 Iterative process of GSS work

The part of the curriculum treated in this subsection is perhaps the most important one. HPC has traditionally been applied in fields such as physics, chemistry, engineering, etc., that is, in natural sciences or related fields. In contrast to the natural sciences, GSS is not an established science yet. Hence a lot of unclear points exist. It is hard to say in advance, for instance, which problems are suited to be investigated. One result of investigating a system is, so to say, whether there are any results that can be delivered by a further investigation of it. There are even GSS-experts (GSSEs) who deny that GSS can make any justified predictions at all. Those experts only use SISs to find new influences and to break the habitual ways of thinking. Another rather obvious point but with great impact is that the mathematization of GSS has not thrived as far as in the natural sciences; this may lead, for instance, to misunderstandings or conceptual deficits. Finally, not all GSS-problems need a clear question to start with or a clear answer to stop.

All this unclarity shows clearly that the task the HPCEs want to support is different from the case of natural sciences in which optimal solutions can be found once and for all. By contrast, investigations in GSS are continous processes leading to more and more clarity.

This ongoing process is full of iterations and loops. First of all, this may pertain to the SIS-internal processes. For instance, the ABM framework or the simulations may require collection of new data. Where exactly these iterations and loops will appear depends on the GS under consideration. In any case assumptions must be made before and during the process which is why these assumptions must be reflected after the simulation at the latest. Which assumptions lead to good predictions? Which ones are relevant? Second, there are external loops as well. If, for instance, no convincing solution to a given problem can be found, it might be the case that the problem itself must be revisited. The problem and its solution coevolve. Even if a suitable solution to a problem can be found, this solution might change the problem in the long run. Bear in mind that the objects of global systems are often humans that can react on insights gained. Finally, there are investigations in which the acceptance of the results and therefore the necessity for further loops may depend on persons external to the investigations and the corresponding initiators.

One example for loops being dependent on persons external to the investigations is politics. A group of politicians might formulate or revise policy objectives and constraints. This begs the question which policy is needed and in which way it can be realised. So, policies and implementation plans must be developed, simulated and evaluated. Of course, this is where HPC-expertise is needed. Regarding evaluation, however, the group of politicians has the final say. If the outcomes are not as desired, new policies together with implementation plans must be developed and again be tested. Only if the group is satisfied with one solution, the respective policy along with its implementation plan can be accepted on the next higher level, for instance by the voting of citizens. In case of rejection, new objectives and constraints must be generated. This is the second loop, which this time is independent of the scientists or initiators. In general, the HPCEs must be willing to accept that there are important parts of the GSS-processes they cannot influence. So, the HPCEs must be prepared that external loops might be necessary.

As any interdisciplinary cooperation, the combination between GSS and HPC takes an effort. Supporting a new field involves getting used to different processes and styles of work, especially if the new field is less closely related to previous ones than these were among each other. The topic of interdisciplinarity, the development of a common language, and many related aspects are treated in the Annex to D2.4.

3.2 Best practices for education and traning

A few “offline” training activities relating to CoeGSS were carried out by different partners, that allowed to gather feedback for identifying best practices for education and training on relevant elements of the CoeGSS workflow.

3.2.1 DSLs of Mathematics

Patrik Jansson and Cezar Ionescu (Chalmers) have worked together with a pedagogical project at Chalmers to develop and improve a BSc level university course on "Domain Specific Languages of Mathematics (DSLsofMath)". All the course material is available online at https://github.com/DSLsofMath/DSLsofMath. The course contents (including exercises) are collected in lecture notes (currently 108 pages, and growing).

The course was developed in response to difficulties faced by third-year computer science students in learning and applying classical mathematics (mainly real and complex analysis). The main idea is to encourage the students to approach mathematical domains from a functional programming perspective: to identify the main functions and types involved and, when necessary, to introduce new abstractions; to give calculational proofs; to pay attention to the syntax of the mathematical expressions; and, finally, to organize the resulting functions and types in domain-specific languages.

CoeGSS is aimed at HPC-enabled policy making in the context of Global Systems Science (GSS). Any GSS model is bound to involve different scientific fields and often the common specification language is "mathematics". But to be "computer-aided" means we also need implementation languages, tools and methods. There often is a large gap between a mathematical model and its implementation. As an example, network diffusion and contagion problems can be specified mathematically in a few equations, but a corresponding HPC implementation can be thousands of lines of C code. One CoeGSS project methodology to bridge this gap is to use domain specific languages (DSLs), that are mentioned above as part of the proposed curriculum for future HPC technology leaders related to GSS. In particular, this means that DSLs of mathematics would be very useful in this curriculum. At the time of the CoeGSS project application (in 2014) there was no such course in the Chalmers curriculum and we therefore handed in a local Chalmers application for a pedagogical project in parallell with CoeGSS. Both applications were granted and by the end of CoeGSS the third instance of the course has been carried out.

Simplified versions of research results from CoeGSS (also reported in publications by Jansson and Ionescu) have been integrated in the course; the closest matches are the lectures on using linear algebra for graphs (networks in CoeGSS), and lectures on DSLs for high level modelling (DSLH, DSLP, DSLD in CoeGSS).

3.2.2 Agent-based modelling and synthetic populations

TOP-IX and GCF organised the presence of CoeGSS in a workshop, "Excursus: Agent-based modelling and synthetic populations" on June 25-26, 2018, within the BigDive7 data science training programme that took place from June 18th to July 13th, 2018 in Turin, Italy (https://www.bigdive.eu/becomedatascientist/big-dive-7/).

In two days that alternated presentations and hands-on excercises, Sarah Wolf and Andreas Geiges from the Global Climate Forum had the chance to present the Mobility Transition Model MoTMo developed within the Green Growth pilot study in CoeGSS WP4 (see D4.6) and to use the agent-based modelling framework ABM4py (see D3.8), even while still under development. The sessions spanned an arch from the role of digital decision support for gobal challenges and the particular example of a sustainable mobility transition, via an overview on synthetic information systems – with synthetic populations and networks, ABM, and considerations on the analysis of simulation output – and some theory behind the global systems approach, to the CoeGSS tools MoTMo and ABM4py. In the practical exercises, groups of 2-3 students implemented decision rules for agents into a code skeleton for a simple ABM on innovation diffusion equipped with a simple synthetic population of Italy, with aim of matching data on the diffusion of compact discs taken from the literature. Approaches taken by the students differed from coarse assumptions via implementing a model found in the literature to analysing the distribution of various features of the population for matching the given curve.

Feedback to the course was overall positive. Participants evaluated all sessions as informative, helpful, structured and understandable and found the combination of presentation and exercise sessions balanced. They requested more details about the framework, in particular, its algorithmic structure, and a technical deepening in general. Based on this feedback, the online version of this course includes more technical detail and encourages to build one's own ABM from scratch, in addition to the option of starting from a skeleton of code.

3.2.3 Online training

As the Big Dive Workshop matched the outline for the Massive Open Online Course (MOOC) on synthetic populations presented in D6.6, the opportunity of turning course material into an online course was exploited by introducing the course "Synthetic Information Systems – the CoeGSS Approach" in the portal.

After an introduction and an outline what can be expected from the course, the first lesson explains why social systems are hard to predict. Using the example of a sustainable mobility transition, elements that combine into potential evolutions of global systems are pointed out and illustrated for the example of the "car centered global system". Potential evolutions, and scenarios are dicussed in particular with a view to transformations. A landscape of theoretical elements to draw on for studying these is presented.

The second lesson focuses on agent-based models for investigating possible system behaviours and the third one introduces synthetic populations, which represent a real-world population in a model, together with synthetic networks, which represent the connections between agents. Both lessons provide an overview what these elements of synthetic information systems are, what steps are needed to develop and use them, and a list of references for participants interested in deepening the subjects.

These overview lessons are accompanied by three hands-on experiments that guide participants towards building their own synthetic population, network, and agent-based model using tools developed by CoeGSS.

The fifth and last lesson then addresses the topic of decision support for global challenges. The presentation shows how synthetic information systems can and need to be embedded into a dialogue with stakehoders that want to use the information but can in turn also help to improve models used, as it were in a balance of model work and stakeholder involvement.

As a dissemination activity beyond the project duration, in the last weeks of the project we explored whether and how the material can also be turned into a follow-up MOOC of the GSS MOOC by Jeff Johnson in the FutureLearn platform, to benefit from the larger user numbers that this course already attracts.

4 Standardisation

As already presented in the previous deliverables D6.5 and D6.6, the standardisation process en-detail is defined as shown in Figure 1:

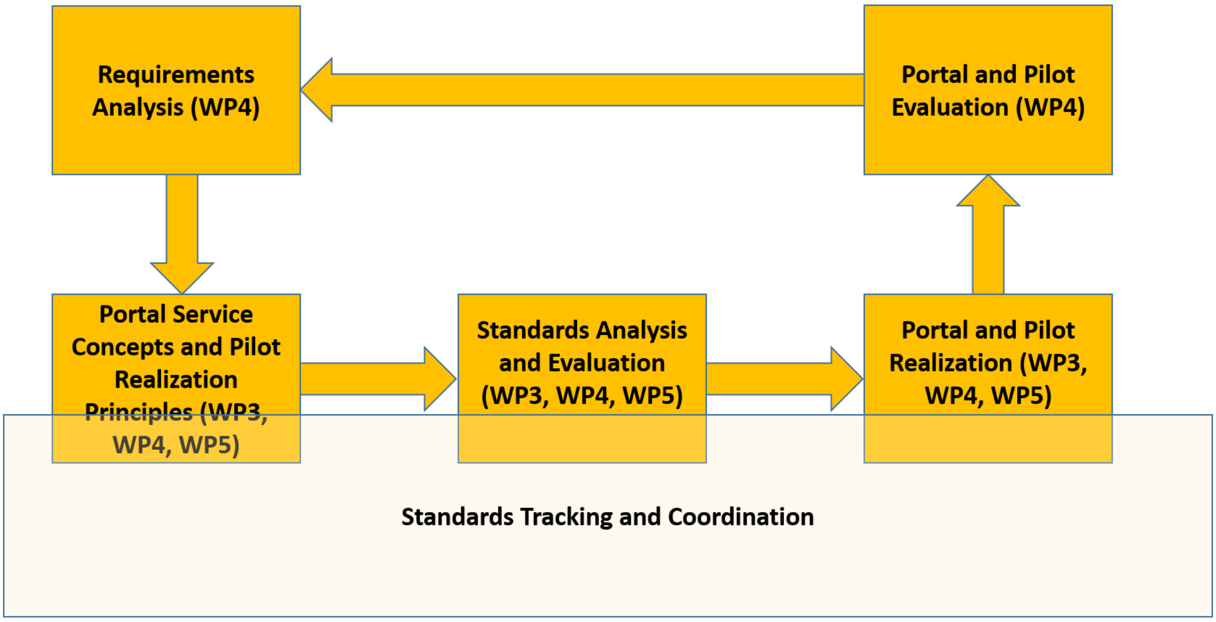

Figure 1: Task 6.4 Tracking of Standards Usage and Potential Contributions

These deliverables also clarified possible standardisation paths:

- the generation of new standards as well as
- the contribution to already existing standards.

At M36 of the project, and thus the end of CoeGSS as a funded EC-Activity, the continuation of the move from standards influencing to standards monitoring has manifested.

As detailed in D6.5, it was not reasonable to create any new standard within project, as the GSS community can be rather seen as a consumer, but not developer of standards. Thus, the main focus of the project was on contributions to standards (where possible, either as reference implementations or as input to the standards descriptions) and furthermore the monitoring of standard candidates.

This was also reflected by a decrease of efforts assigned to this activity; efforts were re-allocated to support activites with stronger impact, however, no complete lock-down of this task was performed, to ensure an interface to the standardisation world.

The following list reprises candidate standardisation activities for the GSS and HPC communities, which are still valid until the end of the project and beyond:

- Global Systems Science

For Global Systems Science in the context of High Performance Computing, there are basically no standards and standardisation bodies available. Nonetheless, after a careful evaluation of the potentials, the project did decide not to put in major efforts to create such standards or even bodies from scratch as the effort would have superseded the benefits.

- High Performance Computing / High Performance Data Analytics / Agent-Based Modelling

For HPC and HPDA, there are various standardization activities ongoing, which are highlighted in the list below.

- NESSI, http://www.nessi-europe.eu
- ETSI, http://www.etsi.org
- CEN, https://www.cen.eu
- SNIA, http://www.snia.org
- ITU, http://www.itu.int
- OGF, https://www.ogf.org
- NIST, https://www.nist.gov

In the diverse activities of the project, results of those bodies were applied where possible.

A major activity within CoeGSS was the definition of a de-facto standard for test benchmarks, analyzing the applicability of HPC architectures for GSS-related applications.

A natural question which arrives at the stage of proving the efficiency of that approach is which HPC architecture is the best one and whether it should be analysed only for efficiency (shortest runtime) or also other factors.

This was the reason to find out a best test (benchmark) which would take into account all the necessary criteria and give the answer which HPC system would be the most appropriate to run GSS applications.

The proposition of the **benchmark** was defined in two project deliverables and finally includes applications used inside the GSS community and compares these applications across the new available HPC architectures. It takes into account the end user requirements in terms of application efficiency – measured by speed and scalability. Moreover, authors decided to introduce two additional metrics for ranking of the tested architectures:

- Energy efficiency calculated as a product of walltimes and TDP products which scales and binds the achieved timing results by processors by the theoretical heat generated during the tests and/or the energy consumed by processors.
- Cost efficiency using scaled timing results by the cores price falling on the given number of cores.

Purposeful distinction of the above metrics may provide valuable data for possible GSS cluster owners and users. In particular, they may use the data to answer many questions at the time of collection or investments planning, such as which processor architecture suits better or is cheaper for use with GSS applications. They may also anticipate and estimate the cost of the energy used by the system and, thus, provide a balance between system performance and its cost. Of course, results presented are purely based on the data provided by processor vendors (TDPs, costs) and do not include other important factors and obvious costs related to HPC infrastructure components.

The benchmark itself is combined by 7 applications which reflects, according to the current knowledge, the best behaviour and requirements of CoeGSS and GSS in general:

- OpenSWPC using swpc_3d.x application with grid size 1000 x 875 x 200 and MPI partitioning relevant to actual tested MPI ranks.
- IPF application with input file input_40k_3200M and test iterations equals to 5. ROWBLOCK and COLBLOCK parameters were set to 4
- GG (Green Growth) application using two maps with different resolution: European map (EU map) with the size of 640x680 and worldwide map (World map) with the size 8640x3432 processes
- CMAQ/CCTM using CMAQv5.2_Benchmark_SingleDay_Input_09_12_2017.tar.gz and CMAQv5.2_Benchmark_SingleDay_Output_09_12_2017.tar.gz datasets downloaded from CMAQ project website
- ABMS with two test cases using layer shape of 64x64 and 128x128
- CM1 with nx = 200 grid points in the x direction, ny = 256 grid points in the y directions and nz = 40 grid points in the z direction
- Hurricane Weather Research and Forecasting (HWRF) model with historical dataset of Hurricane Katrina based on three days 28.08.2005 – 30.08.2015, interval_seconds = 10800, grid dx = 0.193384, dy = 0.191231.

A detailed description of the designed benchmark can be found in deliverables D5.7 and D5.8.

5 Dissemination and Communication

5.1 Dissemination plan

The dissemination plan presented in D6.5 defined forms and types of dissemination for the project by audience, purposes, methods, message and timing. Progress along this plan was documented in D6.6. Dissemination activities such as newsletters and website have further been carried out, see section 5.2 below.

Following the dissemination plan, the third project year moved towards stakeholder dialogues to obtain feedback on the pilot work (see also D2.4). Additionally, industrial, academic and societal awareness of CoeGSS was encouraged through the presentation of results in the events listed in Table 1 below.

5.2 Report on dissemination and communication activities in the third year

5.2.1 Newsletter

During the third project year, the fourth, fifth and sixth newsletter have been sent out. They can be found at http://coegss.eu/events-news/#email.

The project's fourth newsletter was published on October 09, 2017. It was a short edition that announced the second open conference of the project, described above in Section 2.1.1.

On February 20, 2018, the fifth newsletter was published. It covered the relation of HPC-GSS and Big Data and also the potential of HPDA in GSS. The partners Atos SE and ISI Foundation were presented and the Green Growth pilot study showed some results.

The sixth and last newsletter, sent out at the project's closing, discusses the completion of the project, puts a focus on the task about interdisciplinary cooperation, presents the remaining partners, and introduces the portal.

5.2.2 Presentations at workshops and conferences

Table 1 lists presentations at workshops and conferences by partners throughout the third project year.

Date	Partner	Occasion	Aspects, comments
2016-09	Chalmers	International Workshop on Type-Driven Development	Sandberg Eriksson, A. and Jansson, P. (2016) An Agda formalisation of the transitive closure of block matrices. Related to WP3, task T3.4 on DSLs and IPF.

2017-09	UP	Internal workshop	Interval arithmetic for validation Related to WP3, task T3.2 Uncertainty.
2017-10	GCF	Poster presentation at Electric Vehicle Symposium, Stuttgart	Green growth pilot work
2017-10	GCF	Presentation at CPGC conference, Lucca	Green Growth pilot work
2017-11	GCF	Organisation of side event and presentation at the side event at COP23	Green growth thematic and pilot work
2017-11	HLRS	Automated CoeGSS presentation at the SC17	Project presentation
2017-12	Chalmers	Keynote at "GSS in H2020 and beyond" in Brussels	GSS applications and challenges, http://www.cse.chalmers.se/~patrikj/talks/Jansson_GSS_Applications_keynote.pdf Related to all CoeGSS WPs.
2018-01	Chalmers	University course on "Domain Specific Languages of Mathematics"	DSLs, mathematical modelling, etc. Closely related to WP3 task T3.4.
2018-01	HLRS	EuroHPC	Project representation
2018-02	HLRS	IASS Workshop	Project presentation and ABMs discussion
2018-06	Chalmers	Presentation at The Symposium on Trends in Functional Programming (TFP)	Type Safe Interpreters for Free – A method interfacing a functional DSL with an untyped frontend in a safe manner. http://www.cse.chalmers.se/~myreen/tfp2018/slides/Patrik_Jansson.pdf Related to WP3, task T3.4 on DSLs.
2018-06	Chalmers	Presentation at The Symposium on Trends in Functional Programming (TFP)	Examples and Results from a BSc-level Course on Domain Specific Languages of Mathematics. http://www.cse.chalmers.se/~patrikj/talks/TFPIE_2018_Solrun_DSLsofMath.pdf Related to WP3, task T3.4 on DSLs.
2018-06	HLRS	ISC 2018	Project representation

2018-06	HLRS	bwHPC-S5	Project presentation
2018-06	GCF	Training workshop	ABM, synthetic populations, CoeGSS approach to decision support in view of global challenges, GG pilot, MoTMo. Related to WP4.
2018-06	GCF	Stakeholder workshop	CoeGSS approach and GG pilot, MoTMo
2018-07	UP	Talk at the meeting of IFIP Working Group 2.1 on Algorithmic Languages and Calculi	"Clean up monad" Related to WP3, task T3.2 on Uncertainty.
2018-09	IMT	Four talks at the CCS2018, Thessaloniki, Greece	Presentation of the developments on networks studies developed by IMT in CoeGSS. Related to WP3.
2018-09	GCF	Invited talk at CCS2018, Thessaloniki, Greece	Presentation on Green Growth as a complex global challenge, MoTMo as example
2018-09	UP	Talk at the NPFL 2018 Numerical Programming in Functional Languages workshop	"Error analysis almost for free" (Report of UP contributions to T3.2)

Table 1 – Presentations by partners at conferences and workshops

5.2.3 Publications

Table 2 lists publications by partners that relate to CoeGSS from the third project year.

Partner	Type of publication	Complete bibliographical description with DOI, ISBN/eSSN	Peer review	Open access	Presented aspect/topic w.r.t. CoeGSS, target audience
Chalmers	Article in Journal	"The impact of uncertainty on optimal emission policies", N. Botta, P. Jansson, C. Ionescu, Earth System Dynamics. https://www.earth-syst-	Yes	Yes	Global Systems Science: computer aided policy advice with verified implementation. Target audience: the scientific community.

Partner	Type of publication	Complete bibliographical description with DOI, ISBN/eSSN	Peer review	Open access	Presented aspect/topic w.r.t. CoeGSS, target audience
		dynam-discuss.net/esd-2017-86/			The framework described in the previous publications is applied for specifying and solving sequential decision problems to study the impact of three kinds of uncertainties on optimal emission policies in a stylised sequential emission problem. The structure of this problem is similar to those addressed by the CoeGSS pilots. For example, it is similar to the study of effects of incentives and political regulations related to electrical vehicles, currently conducted within the Green Growth pilot.
IMT	Article on arXiv	J. van Lidth de Jeude, R. Di Clemente, G. Caldarelli, F. Saracco, T. Squartini, “Reconstructing mesoscale network structures”, https://arxiv.org/abs/1805.06005	No	Yes	Application of the network reconstruction strategy to directed networks with mesoscale structures
IMT	Article on arXiv	C. Becatti, G. Caldarelli and F. Saracco, “Entropy-based randomisation of rating networks”, https://arxiv.org/abs/1805.00717	No	Yes	Application of the network reconstruction strategy to categorical, signed and review networks
GCF	Article on SSRN	J. Mielke, A. Geiges, “Model-Stakeholder Interactions for a Sustainable Mobility Transition”	No	Yes	Describes how the model from the Green Growth pilot can be used for iterative decision support dialogues

Partner	Type of publication	Complete bibliographical description with DOI, ISBN/eSSN	Peer review	Open access	Presented aspect/topic w.r.t. CoeGSS, target audience
		https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3245159			
GCF	Working paper	G. Steudle, S. Wolf, J. Mielke, C. Jaeger, "Green Growth Mechanics – The Building Blocks" https://globalclimateforum.org/wp-content/uploads/2018/01/GCF_WorkingPaper1-2018.pdf	No	Yes	Foundations of Modelling Green Growth, background for the pilot on that topic
GCF	conference paper	S. Wolf, A. Geiges, S. Fürst, G. Steudle, J. von Postel, C. Jaeger, "Electric Mobility in View of Green Growth", presented at Electric Vehicle Symposium, Oct 2017	No	No	The paper presents initial work of the Green Growth pilot. (In order to guarantee open access, it will be published as a GCF working paper.)

Table 2 – Publications by partners

5.2.4 News Items, Press Releases, Web Presence

Table 3 presents web-based communication items on CoeGSS from partners or external sources during the third project year.

Partner / external	*Activity*	Aspect presented
ALL	*Newsletter No. 4*	2nd Lucca Conference
ATOS	*Internal communication and publicity (intranet, internal bulletin, etc).*	Webinar about the CoeGSS project

ATOS	*Internal communication and publicity (intranet, internal bulletin, etc.).*	CoeGSS progress and potential collaboration with MSO4SC
ALL	*Newsletter No. 5*	HPC-GSS and Big Data, GSS and Data Analytics, Green Growth Pilot
ALL	*Newsletter No. 6*	Project closure, interdisciplinary communication, portal

Table 3 – News items, press releases and similar communication items

6 Conclusion

Over the final year of CoeGSS, activities of all tasks in WP6 were finalised. Highlights reported on in this deliverable include the "International Conference: Computing Power for Global Challenges", that intensified external contacts of CoeGSS and created new ones, and that will remain visible via interviews with speakers and recorded sessions on the website; they also include a curriculum and training material, benchmarking, and a list of publications, that will be available and can be used beyond the project's lifetime (e.g., via the CoeGSS website and portal that will remain active for at least a year after the project finishes). The experience gained in CoeGSS on bringing together HPC and GSS can thus be fruitful for future endeavours with a similar aim.

Throughout the duration of the project, CoeGSS disseminated and communicated the new combination into the two fields that it combined and to adjacent areas, for example, the community of complex systems science. Presentations at workshops and conferences increased from twelve and eleven in the first two years up to twenty in the last project year. A total of over 20 papers was produced about CoeGSS work. In part, these were accepted by scientific journals, others are available as working papers on internet platforms (e.g., arxiv) and will be submitted to journals beyond the project end. Every year the consortium (co-) organized several activities, adding up to nine events as well as four training programmes.

As in previous years, community building activities in the online world and the respective numbers for the overall project duration are reported on by the companion deliverable, D6.4, on the Community Building task. To present the numbers in one place for a concise overview, D6.4 (that is led by the workpackage leader TOP-IX) also contains the total numbers for the other KPIs of this workpackage.

As a closing remark on HPC-GSS community building in the "offline world", let us mention the proposal for a follow-up of the CoeGSS project (EDGE: Exascale, Data, and Global Evolutions, presented by prof. Patrik Jansson, Chalmers). The project would have brought together seven partners from CoeGSS with 12 new partners, including four players from the HPC world, as well as new GSS partners. While not successful as a project, this activity provides proof that community building by CoeGSS has been successful.

7 References

D2.4: CoeGSS Deliverable D2.4; SUSTAINABILITY REPORT; to be released.

D3.8: CoeGSS Deliverable D3.8; REPORT OF FRAMEWORK FOR PROTOTYPING OF PARALLEL AGENT-BASED MODELLING SYSTEMS; delivered to the European Commission; Andreas Geiges (editor), Sarah Wolf, Gesine Steudle, Steffen Fürst

D4.6: CoeGSS Deliverable D4.6; THIRD STATUS REPORT OF THE PILOTS; delivered to the European Commission; Sarah Wolf (editor), Margaret Edwards, Steffen Fürst, Andreas Geiges, Jette von Postel, Michele Tizzoni, Enrico Ubaldi

D5.7: CoeGSS Deliverable D5.7; FIRST REPORT ON PROVIDED TESTBED COMPONENTS FOR RUNNING SERVICES AND PILOTS; delivered to the European Commission; Norbert Meyer (editor), Steffen Fürst, Michael Gienger, Sergiy Gogolenko, Radosław Januszewski, Damian Kaliszan, Michał Pałka, Sebastian Petruczynik, Enrico Ubaldi

D5.8: CoeGSS Deliverable D5.8; SECOND REPORT ON PROVIDED TESTBED COMPONENTS FOR RUNNING SERVICES AND PILOTS; Norbert Meyer (editor), Michael Gienger, Sergiy Gogolenko, Andreas Geiges, Damian Kaliszan, Sebastian Petruczynik, Radosław Januszewski, Paweł Wolniewicz

D6.4: CoeGSS Deliverable D6.4; THIRD ANNUAL REPORT ON COMMUNITY BUILDING; to be released

D6.5: CoeGSS Deliverable D6.5; FIRST ANNUAL REPORT ON TRAINING, STANDARDISATION, COLLABORATION, DISSEMINATION AND COMMUNICATION; delivered to the European Commission; Sarah Wolf (editor), Andreas Geiges, Michael Gienger, Fabio Saracco

D6.6: CoeGSS Deliverable D6.6; SECOND ANNUAL REPORT ON TRAINING, STANDARDISATION, COLLABORATION, DISSEMINATION AND COMMUNICATION; delivered to the European Commission; Sarah Wolf (editor), Michael Gienger, Jette von Postel, Fabio Saracco