CoeGSS

Centre of excellence

D6.6 – Second 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	30.09.2017
Actual date of submission	30.09.2017
Dissemination level	Public
Nature	Report
Version	2 (after internal review)
Work Package	6
Lead beneficiary	GCF
Responsible scientist/administrator	Sarah Wolf
Contributors	Michael Gienger, Fabio Saracco, Jette von Postel

Internal reviewers	Alexandra Gens, Bastian Koller
Keywords	Training, Standardisation, Event Management, Collaboration, Dissemination, Communication
Total number of pages:	27

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		09/2017
Second Version	for submission		09/2017
Reviewed by	Alexandra Gens Bastian Koller	DIA HLRS	09/2017 09/2017
Approved by	ECM	UP, HLRS	09/2017

Abstract

This meta-deliverable presents the state of work in several tasks of WP6 "Awareness Creation and Community Support" at the end of the second 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 activities and achievements include the organisation of two open conferences, elements of a GSS-focused curriculum within HPC, and a list of dissemination and communication activities carried out.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ABM	Agent-Based Model
API	Application Programming Interface
ATOS	Atos Spain SA
BioExcel	Centre of Excellence for Computational Biomolecular Research
CoSMo	The CoSMo Company, Lyon
CS-DC	Complex Systems Digital Campus
DG Connect	Directorate General for Communications Networks, Content & Technology
EPCC	Edinburgh Parallel Computing Centre
ExDCI	European Extreme Data & Computing Initiative
GCF	Global Climate Forum
GDAL	Geospatial Data Abstraction Library
GIS	Geographic Information System
GSS	Global Systems Science
HLRS	High Performance Computing Center Stuttgart
HPC	High-Performance Computing
ICSP2017	International Conference on Synthetic Populations 2017
IMT	School for Advanced Studies, Lucca
IPF	Iterative Proportional Fitting
IPU	Iterative Proportional Updating
ISI	Institute for Scientific Interchange, Turin
ISTI-CNR	Institute of Information Science and Technology of the Italian National Research Council, Pisa
MaX	Materials Designs at the Exascale
MIDAS	Models of Infectious Disease Agent Study
MOOC	Massive Open Online Course
MPI	Message Passing Interface
PSNC	Poznan Supercomputing & Networking Center
QGIS	Quantum-GIS
SIS	Synthetic Information System
SPEW	Synthetic Populations and Ecosystems of the World

1 Introduction

This deliverable presents the second 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

A section is dedicated to each of these tasks below. As mentioned in D6.5, there is a close relation between all tasks in WP6. To avoid overlap, D6.3 "Second Annual Report on Community Building", which describes progress made in T6.1 is referenced wherever needed.

To make this deliverable a standalone document, the following paragraphs very briefly introduce CoeGSS. They can be skipped by readers familiar with the project and previous deliverables. The Centre of Excellence for Global Systems Science brings together High Performance Computing and Global Systems Science. Three pilot studies – Health Habits, Green Growth, and Global Urbanization – address a global challenge each with the help of HPC-based modelling and simulation: smoking habits, the diffusion of electric mobility in the global car fleet, and two-way relations between transport infrastructure and real-estate pricing, respectively. They develop synthetic information systems (SISs), to represent the real-world system under study on the computer and run simulations for exploring, as in a virtual laboratory, possible scenarios of the system's future evolution. In particular, such a system contains a synthetic population and an agent-based model. The former is a set of virtual agents that, for relevant characteristics, statistically match the corresponding distributions found in the real-world population represented. The latter is a simulation model, which implements interactions of agents. Running the agent-based model initialised with the synthetic population that is, repeatedly carrying out the virtual agents' interactions and then analysing the thus created simulation output, the synthetic information system helps to understand potential system evolutions and helps to assess possible consequences of decisions.

2 Event Management and Collaboration

2.1 Events organised

2.1.1 International Conference

The first International Conference on Synthetic Populations (ICSP2017) took place in Lucca, 22-23 February, 2017. It aimed at making the communities of Global System Science and High Performance Computing scientists meet and contribute to the ongoing discussion on the natural common ground of the studies on synthetic populations. With over 65 participants, it consolidated the emerging community of scientists and practitioners in this area and fostered the discussion from both the modelling and the computer science point of view. World leaders in working with synthetic populations were attracted, and participants from CoeGSS mingled with specialists from other centres of excellence.

A session on Centres of Excellence gave an overview of ExDCI (European Extreme Data & Computing Initiative), and the Centres of Excellence BioExcel (The Centre of Excellence for Computational Biomolecular Research) and MaX (Materials Designs at the Exascale).

The CoeGSS project was represented by several partners in different sessions. Carlo Jaeger (Potsdam University) and Bastian Koller (HLRS) introduced the issue of using HPC for addressing global challenges from a GSS and an HPC perspective. In the Health and Epidemics Studies session, Alessandro Vespignani (ISI) talked about contagion processes in complex societies, and Luca Rossi (ISI) presented work on the combined execution of an agent-based model and a metapopulation, both relating to T4.1, the Health Habits pilot. Sarah Wolf (GCF) opened the Green Growth and Sustainable Mobility session with an insight to electric mobility in view of green growth, presenting the state of work in T4.2, the Green Growth pilot, at that moment. Completing the Smart Cities and Global Urbanisation session, Margaret Edwards (CoSMo) promoted T4.3, the Global Urbanization pilot in her talk on city synthetic populations: unfolding some problematics of complexity simulation by studying the influence of data and agent granularity.

Further sessions addressed Agent-Based Models and Information Systems, Financial Systems, and Policy Informatics, featuring in total 18 keynote speakers from ten different countries including Stefano Battiston (University of Zurich), Chris Barrett (Virginia Tech), Madhav Marathe (Virginia Tech), Jeff Johnson (The Open University), Stefan Thurner (Medical University of Vienna) and Guido Caldarelli (IMT). Thanks to a cooperation with the Simpol project, the conference presentation could be videotaped and streamed. The videos and interviews with the keynote speakers can be found on the conference website (https://icspconference.wordpress.com/abstract-submission-2/). Together with the organisational effort of IMT and the wonderful location and environment, the event led to an exchange of experiences and ideas in an unusually creative atmosphere. The conference was

so fruitful that the participants thought it would be great to have a next open conference on the same premises (see Section 2.2).

2.1.2 Informal workshop

Suggested by the project officer Carlos Morais Pires, CoeGSS organized an informal workshop on bringing together HPC and GSS, in conjunction with the project review at the end of May, 2017. An overview of the European HPC strategy was given by Andrea Feltrin (DG Connect, European Commission), and keynote speeches on GSS and HPC, respectively, were delivered by Sander van der Leeuw (Arizona State University) and Mark Parsons (EPCC, University of Edinburgh). Discussions between experts on HPC and GSS showed a mutual interest in how the two fields can create something that is more than the sum of the parts by working together.

On the one hand, with large amounts of information, processed in great detail, becoming available, there is a potential for GSS to improve our understanding of societal and environmental phenomena, for example in moving from statistical, sample-based analysis of global systems' dynamics to modelling at the level of the individual. In the longer term, this may require extremely large computing resources.

On the other hand, extreme levels of parallelism have been and continue to be a software challenge, especially in approaching exascale computing. The Centres of Excellence are expected to support co-design, evaluation and validation work for the future HPC Programme of the EU, for example with the help of extreme scale demonstrators. While some of the other Centres of Excellence are already core-users of petascale HPC systems, the GSS community is only beginning to consider the use of HPC systems. Rather than the development of mature GSS-HPC applications, which cannot yet be expected, two tasks are therefore relevant for CoeGSS: developing smaller scale demonstrators, and outlining the potential of longer-term developments in this direction. To quote Mark Parsons: without the baggage of 50 years of development, the HPC modelling and simulation community and the Global Systems Science community should be working together to define Grand Challenges that require Exascale computing.

2.2 Events planned

Following up on ICSP2017 (see Section 2.1.1), CoeGSS is currently organising the International Conference: Computing Power for Global Challenges (see http://cpgc.coegss.eu). This event shall bring together HPC and GSS scientists with practitioners working on four global challenges:

- Developing a sustainable and resilient global financial system
- Addressing the daunting risks of pandemics
- Transforming the fossil-fuel based global mobility system

- Creating forms of democracy adequate to the age of digitalization

In panel discussions and workshops, the event shall explore how GSS and HPC can join forces in further development of large scale simulation modelling for addressing societal challenges. For the four examples, it shall investigate GSS applications that show a need for HPC support to enable them to advance beyond the current state. It shall further look into directions of research and development at the intersection between HPC and GSS that can accommodate GSS structures in HPC-efficient simulation models and can lead to HPC-efficient simulation analysis methods for these. With the help of inputs from practitioners, the discussions shall be grounded in the practical dimensions of addressing global challenges, allowing to focus on the type of questions decision makers have and to be informed by their expertise.

2.3 Collaboration

The networks of CoeGSS partners cover a large spectrum of potential stakeholders and collaborators that were presented in D6.5, and shall therefore not be repeated here. Additional collaborations include

- a project together with the Division of Building Technology at Chalmers, the Energy in Urban Development Chalmers area of advance and the Network Dynamics and Simulation Science Laboratory at Virginia Tech. This project investigates energy consumption in buildings, using synthetic populations generated at Virginia Tech. Devdatt Dubashi and Oskar Allerbo from Chalmers were involved in both this project and CoeGSS over the second project year.
- a research collaboration between IMT researchers and Luca Pappalardo (ISTI-CNR Pisa) and Dino Pedreschi (University of Pisa) on the reconstruction and validation of retail market networks,
- discussions between ATOS and the Tango project on the role of HPC in niche and new applications,
- an exchange between GCF and the SPEW group at Carnegie Mellon University, that led to a tutorial on synthetic populations by Shannon Gallagher from this group (see Section 3.3), co-organized with the International Conference on Synthetic Populations.

The cross-linking of websites, which can be considered an activity in community building and as part of collaboration, is described in D6.3.

3 Training

Throughout the second project year, several activities were carried out in the training task to build on what was described in D6.5 in working towards the definition of a curriculum for future HPC technology leaders related to GSS needs (see Section 3.1) and the identification of best practices for education and training (Sections 3.2 - 3.3). Further, CoeGSS plans to participate in the creation of an online course on synthetic populations, as an activity at the intersection between training and dissemination (Section 3.5). The website with training information (http://coegss.eu/resources/#training) is still running.

3.1 Requirement analysis

While D6.5 focused on introducing GSS to HPC experts and vice versa, during the second year, competencies needed, on the one hand for building a GSS synthetic information system and on the other hand for HPC experts in view of GSS needs, were collected. An overview of competencies relating to SISs is given in the following table (for details on synthetic information systems, see D4.1 and D4.4). These include GSS competencies, such as agent-based modelling skills, as well as HPC competencies, such as practical knowledge on how to eliminate bottlenecks in a parallel implementation of an ABM. The information collected here provides some entry points for what CoeGSS can offer to GSS modellers (see also D2.2): for example, an HPC-ABM framework allows modellers to implement models for use on HPC without needing to learn the competencies required for writing parallel code.

At the same time, the list provides points an HPC expert should be aware of in terms of GSS needs. Without having to become GSS experts themselves, HPC experts will need a passive understanding of steps and competencies needed in GSS model development processes.

Task	**Competency requirements**		
	Practical Knowledge	**Theoretical Knowledge**	**Skills**
Conceptual model			
define agents, networks, environment	knowledge of the system	knowledge of useful theory to be applied	agent-based modelling
Data			
collect	data sources available, data licencing; how to establish access to data and micro data	data conversion	collect and store data in useful mode

Task	Competency requirements		
	Practical Knowledge	**Theoretical Knowledge**	**Skills**
find input data	know where to find suitable input data	understand what data is suitable, know about aggregate data and micro samples	(re)format data
pre-process	API programming	how to assess quality	data analysis
Synthetic populations			
use an existing population	knowledge of existing populations e.g. MIDAS/SPEW		
generate a population	which tools exist and how to use them	maximum likelihood estimation, algorithms e.g. IPF, IPU	software experience in R or Python
add geographic properties	GIS, Shapefiles	coordinate transformations	GDAL, QGIS
extend an existing population	knowledge of existing populations e.g. MIDAS/SPEW, how to use tools	maximum likelihood estimation, algorithms e.g. IPF, IPU	software experience in R or Python
Synthetic networks			
reconstruct relations among agents from limited information	knowledge of the system and of the possible interesting networks of relations for the agents	information theory; graph and network theory; knowledge of existing methods for generating synthetic networks	programming experience in R or Python; expertise in network theory
ABM			
implement	programming languages, commands, libraries, frameworks	graph and network theory, optimization theory	programming
calibrate	data and patterns to calibrate on	computational statistics	statistical analysis
optimize, profile code	C, C++; how to improve code performance and eliminate bottlenecks	parallel and distributed computation, analysis of data and communication flow	performance analysis

Task	Competency requirements		
	Practical Knowledge	**Theoretical Knowledge**	**Skills**
Simulations			
run	C++/Fortran experience	graph partitioning, software development, computing hardware	parallel computing, MPI
analyse	statistical tools (e.g., R, S), possibly coming with script language if pre-processing required (e.g. Python)	theoretical analytical tools (statistics, dynamical studies, AI (e.g., high level pattern recognition)	programming with statistical tools
visualise	visualisation software: going from simulating spatial and time evolution to high level insights (charts of specific indicators, maps of results in the space of parameters)	model analysis needs (key indicators identification, dynamic studies, parameter space exploration, emergence)	visualisation software programming
Presentation			
present results to stakeholders	the system, visualisation	research and modelling done	listen to stakeholders, present work in their language

Table 1 – Competency Requirements

To identify requirements for a curriculum, HPC and scientific computing experts in the project provided information on what they learned about GSS through their work in the project, while GSS experts listed points they considered essential for HPC experts to be aware of for supporting GSS needs. Further, a questionnaire prepared in T2.3 (further information on this will follow in D.2) posed some questions on training needs identified. The remainder of this subsection collects information gained from these activities.

A general insight that was repeatedly acknowledged by CoeGSS participants from all fields was the challenge of merging previously unrelated fields. First of all, this requires an interest in thinking outside of one's disciplinary box and in learning some basics (see also D6.5): what GSS/HPC is, what their benefits are, pros and cons of HPC on a theoretical level, generic GSS ideas and approach. Beyond this, learning by doing in close collaboration between GSS and HPC, potentially involving end users of the models under development, was identified as a requirement. Insights from this process, for example on how to define requirements in a way useful for both sides, can then aid the definition of a curriculum.

Within a synthetic information system for GSS, an element of particular interest for using HPC is an agent-based model and its simulations. A focus where close interactions have taken place

was thus an HPC-ABM framework. Learning about common traits of ABMs as well as spatial and graph based structures in ABMs representing global systems (see D4.2), and learning how to code models with the help of available HPC-frameworks for agent-based modelling and simulation (ABMS), thus exploring the features provided and the limitations of the frameworks were necessary elements for coding an HPC-ABMS framework. While HPC experts working with GSS will not all need to code such a framework, repeatedly occurring structures of GSS ABMs are a point to be considered in the curriculum of interest.

An important and related point is the fact that the model development process in GSS involves an iterative process including model definition, implementation, analysis with visualisation, and refinement. Hence, initial model versions are often simpler than one might expect, and computing power is needed for running and analysing many simulations on the way to more complex models. While such "embarrassingly parallel" simulations are not what HPC systems are designed and constructed for, whether and how HPC can be taken advantage of in these necessary model exploration phases is a point for further common research between HPC and GSS. Due to the absence of macro-level equations for the complex dynamics of global systems, model validation and verification is a challenging task; in fact sensitivity analysis and model calibration in social sciences were mentioned as one of the points HPC experts would like to learn more about. In agent-based models, code often tends to be low-level, and it is not easy to establish whether it really implements the model that the modellers had in their heads, or written down. Therefore modelling using types and functions, type based specifications, and DSLs and generative programming were suggested as useful elements for a curriculum between HPC and GSS.

Similarly, data pre- and post-processing techniques, synthetic population generation (including knowledge about micro samples and statistical sampling techniques), as well as methods and algorithms for network reconstruction, or more generally, techniques of input data generation, were mentioned as points to be included in an HPC-GSS curriculum. These may include recipes for the derivation of network reconstruction algorithms from first principles or a set of ready algorithms that can be applied to broad classes of GSS models.

Further, application purposes should be included as examples in such a curriculum. Learning about the role of matrix inversion in economics, can be a starting point for transferring or adapting HPC tools and codes to GSS applications. Similarly, learning from established epidemics models can help identify modelling needs.

Finally, an HPC-GSS curriculum needs to focus on some areas within HPC that are of particular relevance for working with GSS. Standard knowledge in parallel programming (MPI, Open MP/STL Threads) being assumed, one of the challenges seems to be to know which HPC solutions and libraries to choose for each problem under consideration, due to the amount of libraries existing that have been developed outside a GSS context. Also, skills in adapting such pre-existing solutions, for example for visualisation, to best fit specific GSS needs would need to be trained.

Areas of special interest for supporting GSS, pointed out by HPC experts in CoeGSS, include high-performance graph-parallel analytics (GraphLab/PowerGraph, GraphChi, etc.), high-performance sparse linear algebra, parallel reading and writing of data, machine-independent binary data formats for HPC (HDF5, NetCDF), parallel graph libraries, graph partitioning and parallel data supply, and some advanced topics in C++ programming for HPC (some TMP techniques, compile-time introspection, etc.).

3.2 Feedback from participation in an HPC course

As mentioned in D6.5, HLRS, in cooperation with PATC (PRACE Advanced Training Centre), offered a dedicated 5-days training event in Stuttgart to the members of the CoeGSS project. It took place on October 17-21, 2016 at HLRS Stuttgart (see also the course website https://www.hlrs.de/de/training/2016-10-17-par/). The course taught distributed memory parallelization with the Message Passing Interface MPI and shared memory parallelization with OpenMP, for beginners. It then treated intermediate and advanced topics in parallel programming.

From the consortium, Margaret Edwards (CoSMo), Oskar Allerbo (Chalmers), Francisco Javier Nieto De Santos (ATOS) and Burak Karaboga (ATOS) took part in this course. The participants from the consortium generally found the workshop useful for GSS-scientists as an introduction to HPC. Benefits of this course mentioned were:

- getting an insight into how to think when writing parallel programs and how models need to be built for HPC use
- learning about interesting optimization mechanisms for parallel applications and tools that support developers when developing parallel applications.
- learning about potential data input/output issues
- the mix between theory and practice, and in particular the possibility to run practical exercises on HLRS machines
- not least, the good organisation and the enthusiasm of the instructors about the course and its subject.

From a pilot modeller's perspective, this course pointed out that generic parallelisation of a pre-existing modelling tool (like the CoSMo Simulation Suite, which the Global Urbanisation pilot's model uses) would pose serious challenges, for example because data that would need to be exchanged between processes can hardly be defined at a general level. A point of interest for the CoeGSS pilots, on parallelizing processes based upon complex, possibly evolving networks pointed to the fact that this is not part of the standards taught in such a course; however the instructor considered it an interesting question.

3.3 Training course on synthetic populations

Back-to-back with the International Conference on Synthetic Populations organised in Lucca in February, 2017 (see Section 2.1), CoeGSS organized an internal training workshop on synthetic populations for members of the consortium. In this half-day event, two speakers provided in depth information on activities involved and tools that can be used for the generation of synthetic populations.

First, Shannon Gallagher (Carnegie Mellon University) introduced the R package SPEW (Synthetic Populations and Ecosystems of the World, see http://stat.cmu.edu/~spew), a tool to produce a synthetic population. By creating a synthetic population of Lucca she explained the necessary steps in SPEW: the collection of geographic, population and environmental input data, its harmonization and then the synthesis allowing for different sampling methods. She also gave a short overview on validation and visualization of the population.

Next, Oskar Allerbo (Chalmers) shared insights on the creation of "Synthetic Sweden". He explained the sampling of a static population by, for example, uniform sampling or Iterative Proportional Fitting, and described household grouping within the sampled population as well as the assignment of activities, home locations and activity locations. Last, he introduced a contact network identifying which synthetic agents within the population are at the same place at the same time.

The consortium benefitted from this course by getting a shared overview over synthetic populations, with examples of the kind of data needed, of available sampling algorithms, and of networks that can be constructed between the agents of a synthetic population. After each of the tutorial presentations, question and answer sessions allowed members of the consortium to clarify and to find out how the examples presented relate to certain parts of work in CoeGSS, in particular, to the modelling activities of the pilots.

3.4 GSS online course

The massive open online course (MOOC) "Global Systems Science and Policy: an Introduction" (see https://www.futurelearn.com/courses/global-systems-science/) has been described in D6.5; also throughout the second project year, CoeGSS presence in this course was secured by participating as educators on the course in various presentations (Sarah Wolf (GCF): December 2016, Franziska Schütze (GCF): March, June, September 2017).

This GSS course serves as a good practise example for online trainings in an interdisciplinary field. The course is introductory and easy to follow for persons with diverse backgrounds. The course provides necessary background and combines it with specific use cases. It uses different learning formats, such as videos, reading material, quizzes, experience sharing, discussions and so forth. The interactive discussion format triggers experience sharing across the globe and between different disciplines.

Hence, this activity has helped identify benefits of the MOOC online training format in the context of GSS more generally:

(1) MOOCs can contribute to community building: this GSS course has participants from all over the world with different backgrounds and from different fields of work. By answering questions from these different perspectives and commenting on each other's answers, people widen each other's horizons on the various challenges GSS can and should address. People who previously do not see themselves as involved in GSS, see how their work contributes to this emerging field. On community building at the intersection between HPC and GSS, see D6.3.

(2) The worldwide accessibility of the course allows for wide dissemination. As the educators involved in CoeGSS referred learners to the Centre and informed about its activities and methodology whenever this was an appropriate reaction to their questions or comments, awareness raising was also possible via this course.

(3) The MOOC format that encourages participants to share their knowledge and experience fosters mutual learning. Participants, as well as educators, can benefit from the expertise of other participants. As CoeGSS brings together two previously rather unrelated communities, mutual learning should be encouraged wherever possible.

3.5 Preparation of a synthetic populations MOOC

To reap some of these benefits of the online training format of a MOOC, plans for producing a MOOC on synthetic populations in the context of addressing global challenges have recently been made. Together with Jeff Johnson (The Open University), author and lead educator of the above described GSS Introduction MOOC, some basics have been specified:

- Aim of the course: to give a clear understanding of the theory, implementation and potential uses of synthetic populations in view of addressing global challenges. This MOOC shall provide an introduction; there is potential to deepen certain aspects in a series of further MOOCs. How HPC can play a role in this course, or in more specialised ones that could then follow, will need to be figured out within the consortium of CoeGSS.
- Audience addressed: in the first instance, students and researchers, also government scientists in ministries, or officers working in local authority planning. Further, data scientists for whom this may be another source of (big) data.
- Constraints: the targeted audience will have time constraints, therefore the MOOC shall be restricted to 8-10 hours. In terms of prior knowledge constraints, graduate level study skills but no specialised technical knowledge shall be assumed.
- Structure and content: the different lessons in this MOOC can be:
  - 1. Introduction.

  - 2. Why social systems are hard to predict.
  - 3. Agent-based modelling for investigating possible system behaviours.
  - 4. Synthetic populations – what they are, how they are generated, dynamics, calibration.
  - 5. Hands on experiment with a demonstrator, possibly based on one of the CoeGSS pilot models
  - 6. How to answer policy questions in public administration and business.
- Cooperations: we can take advantage of contacts with people working on the following projects:
  - CS-DC: the Complex Systems Digital Campus is an international network of individuals and institutions working together and sharing resources to promote research and education in complex systems science and in integrative sciences.
  - Da.Re.: The Data Science pathways to re-imagine education project intends to identify and design training programs dedicated to the future of big data professionals, the contents of which will be built on the basis of the actual needs of enterprises.
  - SPEW: Synthetic Populations and Ecosystems of the World provides realistic synthetic ecosystems and supports the generation of synthetic ecosystems for agent-based models.
- Team structure: There will be a small production team, comprising at least one person from each of the cooperating projects. A scientific advisory board shall be invited to help produce the MOOC by commenting on the material.
- Budget: based on previous experience, there will be a need of €5-10k for producing videos, paying copyrights and other cash costs. Parts of this can possibly be contributed from dissemination budgets of projects cooperating in this MOOC.
- Next steps:
  - Clarify the potential of including HPC aspects in this MOOC within CoeGSS.
  - Invite potential cooperating projects, invite people to the production team, invite advisory board members. Form production and presentation teams.
  - Draft a proposal for the MOOC, and present this to Future Learn (that also hosts the introductory GSS MOOC). Circulate the proposal to the advisory board.
  - Begin production.

- Presentation of the MOOC: 4 presentations per year, following the GSS introductory MOOC and advertised by it, seem reasonable. The first presentation should take place in the first half of 2018.

3.6 Outlook

In terms of identifying best practices, Sections 3.2 and 3.3, together with Section 3.4, suggest a combination of online training material, such as a MOOC, for wide outreach, and specific face-to-face events, comprising HPC courses for GSS experts, GSS courses for HPC experts, and workshops that foster mutual learning. Such workshops shall be carried out within the International Conference: Computing Power for Global Challenges that is upcoming on October 24 and 25, 2017. This conference shall also include a focus group session for identifying further elements of an HPC for GSS curriculum.

Training needs to be identified by the consortium (also with the help of the T2.3-questionnaire mentioned in Section 3.1) suggest a whole series of potential training material for an emerging HPC-GSS community beyond the project's lifetime. Some of these could be included in an introductory manner in the MOOC under preparation:

- Introduction to HPC for social scientists
- Introduction on the issue of simulation analysis
- Uncertainty interpretation – to avoid making the wrong decision based on uncertain results. Sequential decision problems, for example.

Others deserve deeper treatments, by themselves, or as follow up material to an introduction:

- Deepening HPC for social scientists (i.e. general concepts, programming paradigms, used languages, resource management, numbers for available memory and what does that mean in terms of concrete formats).
- A tutorial on the use of an HPC-ABM framework that allows modellers to use it in parallel etc. without deep understanding of parallelisation required; how to get up and running HPC simulations
- Possibilities for life-visualisation with the help of HPC and model adaptions (for visual analysis) including security aspects.
- A glossary or a wiki

A priorisation of efforts in the training task for the last project year will be discussed at the upcoming project meeting in October.

4 Standardisation

D6.5 already presented the standardisation process en-detail (cf. Figure 1)

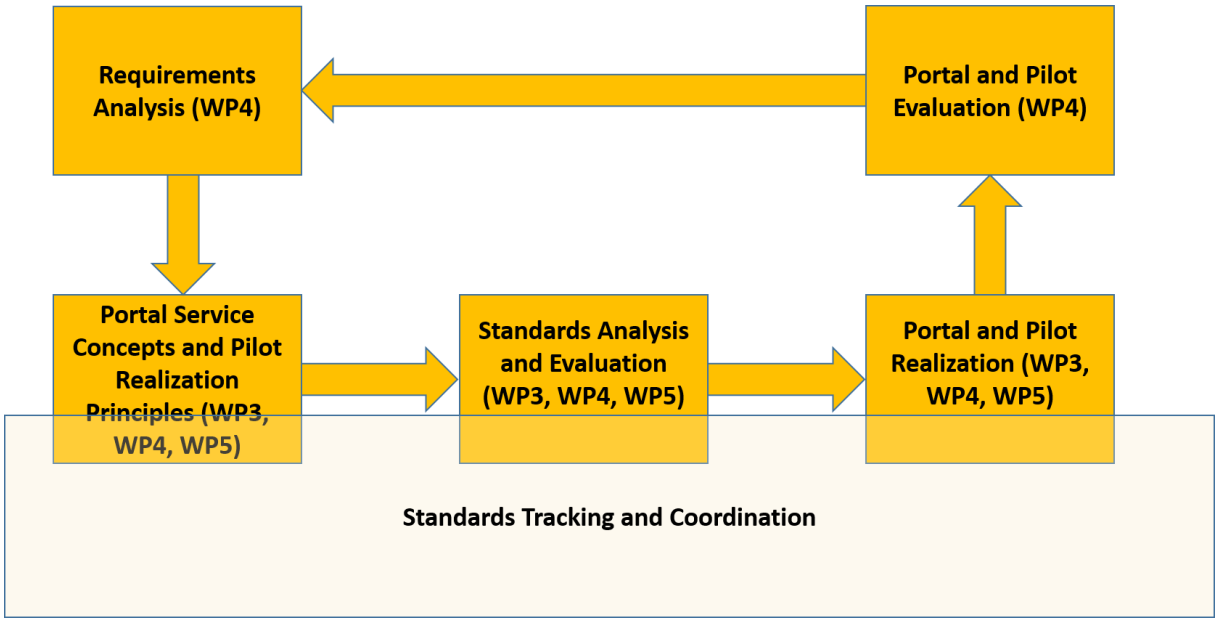

Figure 1: Task 6.4 Tracking of Standards Usage and Potential Contributions

It also clarified possible standardisation paths:

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

Now, with the arrival at M24 of the project, it is obvious that it is not reasonable to create a new standard within the remaining time of the project. This will lead to a main focus on contributions to standards (either as reference implementations or as input to the standards descriptions).

On a technical level, there is still the constant monitoring of standard candidates, however clear contributions could not be performed yet due to the slow progress of technology realizations or simple standard-conform activities. Thus, we are not yet at the stage in the project to really provide maturity of work, which could be of interest for standardisation bodies.

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 so that the list is limited when creating this deliverable. However, all kinds of programming techniques and methods can also be submitted in the area of pure HPC.

  - OpenABM, https://www.openabm.org

- High Performance Computing / High Performance Data Analytics

For HPC and HPDA, there are various standardization activities on-going, 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

5 Dissemination and Communication

The dissemination and communication activities are closely related with the activities for community building, and the event management task.

5.1 Dissemination plan

The dissemination plan presented in D6.5 specified four groups in the audience of CoeGSS dissemination activities: internal, other projects, external stakeholders, and the HPC and GSS communities in general. It defined several dissemination purposes in relation to these groups that have been and will be carried out, including the newsletter and website that were used to raise awareness within the communities, sharing of information within the consortium via internal documents, organisation of events for community building (see Section 2), and presentation of results at conferences or in publications (see Sections 5.2.2 and 5.2.3). According to the dissemination plan, the project focused on awareness raising for a first project phase, adding activities involving personal interactions for the second project phase, among which the organisation of events as the two conferences in Lucca described in Section 2. Moving towards the third phase, for example, the first results of the green growth pilot shall be presented at the Electric Vehicle Symposium in October, 2017; the paper submitted to this event can be found in D4.5 as an appendix.

5.2 Report on dissemination and communication activities in the second year

5.2.1 Newsletter

In the second project year, the second and third project newsletters have been sent out. They can be accessed from http://coegss.eu/events-news/#email.

The second newsletter, published on February 09, 2017, covers the sections introduced in D6.5. News is given on first steps of the Centre and challenges that it is facing. The partners HLRS and GCF are presented, as well as the GSS feature of complex interaction networks and a workflow for data processing in HPC. The Health Habits pilot gave an overview of their work. The newsletter gives an outlook to the CoeGSS-organized International Conference on Synthetic Populations.

The project's third newsletter was published on May 29, 2017. It focuses on the first project conference in Lucca, where a lot of input and new perspectives on synthetic populations were given. The HPC and GSS sections show outcomes of different conference sessions from these two perspectives. The newsletter also presents the partner IMT that hosted the conference as well as PSNC. Also, the Global Urbanisation pilot was introduced.

5.2.2 Presentations at workshops and conferences

Table 2 lists presentations at workshops and conferences by partners throughout the second project year.

Date	Partner	Occasion	Aspects, comments
2016-09	Chalmers	International Workshop on Type-Driven Development	Presentation of Sandberg Eriksson, A. and Jansson, P. (2016) An Agda formalisation of the transitive closure of block matrices
2016-10	IMT	Workshop organisation in collaboration with Tiziana Di Matteo (King's College): "*System Science and Policy*", King's College (London)	Network reconstruction
2016-11	HLRS	Supercomputing Conference 2016	Invited Talk
2016-11	PSNC	Presentation of HPC activities at the Polish Ministry of Science and Higher Education	Scope of presentation included activities in the area of HPC, including the current state in Poland, inter-national co-operation: CoeGSS and PRACE.
2017-02	IMT	Organisation of "ICSP International Conference of Synthetic Populations" in Lucca	The event advertised the activities of CoeGSS and permitted the interaction of our consortium with other centres of excellence present in the EU.
2017-02	several	Presentations about CoeGSS work at ICSP2017	See Section 2.1.1
2017-03	Chalmers	Meeting on Dependently Typed Decision Theories, Potsdam Institute of Climate Impact Research	Relating the SeqDecProb theory described in the paper in Logical Methods in Computer Science (see Table 3) with game theory and infinite horizon decision problems in economics for GSS modelling
2017-03	HLRS	25th Workshop on Sustained Simulation Performance in Sendai, Japan	Project Overview

2017-06	GCF	Presentation at Free University Berlin, Workshop "Mathematics in Digital Humanities and Computational Social Science"	Green Growth, theoretical aspects, underlying structures to represent in models (heterogeneous agents, networks, iterated games)
	HLRS	Collaboration Presentation in Berkeley	CoeGSS project
	HLRS	PRACE Meeting in Juelich	CoeGSS project

Table 2 – Presentations by partners at conferences and workshops

5.2.3 Publications

Table 4 lists publications by partners that relate to CoeGSS from the second 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	Publication in Workshop Proceedings	Adam Sandberg Eriksson and Patrik Jansson, An Agda formalization of the transitive closure of block matrices, Proceedings of the 1st International Workshop on Type-Driven Development, September 2016, ISBN: 978-1-4503-4435-7, available at http://publications.lib.chalmers.se/publication/240988	Yes	Yes	Matrix algebra is central for HPC and this paper contributes to "correct-by-construction" matrix algebra implementation.
IMT	Article in Journal	Matúš Medo and Guilio Cimini, Model-based evaluation of scientific impact, Physical Review E, 94, Sept. 2016, 10.1103/PhysRevE.94.032312	Yes	No	Using bibliometric data artificially generated through a model of citation dynamics calibrated on empirical data, we compare several indicators for the scientific impact of individual researchers. The ABM proposed can be parallelized and it can inspire the pilots.

IMT	Working Paper	Tiziano Squartini, Assaf Almog, Guido Caldarelli, Iman van Lelyveld, Diego Garlaschelli and Giulio Cimini, Enhanced capital-asset pricing model for bipartite financial networks reconstruction, arXiv by Cornell University Library, Oct. 2016,	No	Yes	Development of a new reconstruction method based on statistical mechanics concepts and tailored for bipartite market networks. The method is successfully tested on NASDAQ, NYSE and AMEX filing data, by correctly reproducing the network topology and providing reliable estimates of systemic risk over the market.
Chalmers	Working Paper	Nicola Botta, Patrik Jansson and Cezar Ionescu, Contributions to a computational theory of policy advice and avoidability, Nov. 2016	Yes	Yes	Accepted for publication 2017-09. Computer aided policy advice is at the core of Global Systems Science and this paper presents a specification and a formally verified implementation of Bellman's backward induction algorithm and explains how it can be used for policy advice.
IMT	Article in Journal	Fabio Saracco, Riccardo Di Clemente, Andrea Gabrielli and Tiziano Squartini, Detecting early signs of the 2007-2008 crisis in the world's trade, Scientific Reports, 6, 2016, DOI: 10.1038/srep30286	Yes	Yes	Evolution of the bipartite World Trade Web (WTW) 1995-2010; structural changes in the WTW topology. Relates to network analysis for pilots.
HLRS	Chapter in Book	Bastian Koller, Michael M. Resch, Michael Gienger, High Performance Computing and High Performance Data Analytics – What is the Missing Link?, Sustained Simulation Performance, 2016, ISBN 978-3-319-46734-4	No	No	High Performance Data Analytics High Performance Computing Cognitive Computing

IMT	Article in Journal	Matteo Serri, Guido Caldarelli and Guilio Cimini, How the interbank market becomes systemically dangerous: an agent-based network model of financial distress propagation, Journal Of Network Theory In Finance, 3, March 2017, DOI: 10.21314/JNTF.2017.025, available at https://arxiv.org/abs/1611.04311	No	Yes	This relates to one of the identified potential future applications: contagion in financial networks.
Chalmers	Article in Journal	Nicola Botta, Patrik Jansson, Cezar Ionescu, David R. Christiansen and Edwin Brady, Sequential decision problems, dependent types and generic solutions, Logical Methods in Computer Science, 13, March 2017, DOI: 10.23638/LMCS-13(1:7)2017	Yes	Yes	This is part of a series of papers with the aim to clarify important concept in GSS policy making and to provide implementations with formal proofs of correctness (in Idris). Target audience: the scientific community.
Chalmers	Publication in Workshop Proceedings	Maximilian Algehed and Patrik Jansson, VisPar: Visualising Dataflow Graphs from the Par Monad, Proceedings of FHPC'17, Oxford, UK, September 2017, DOI: 10.1145/3122948.3122953	Yes	Yes	The paper describes a method for visualising the dataflow dependencies of parallel computations using the Haskell library called Par. Target audience: the scientific community

Table 3 – Publications by partners

5.2.4 News Items, Press Releases, Web Presence

Table 5 presents web-based communication items on CoeGSS from partners or external sources during the second project year.

Partner / external	*Activity*	Aspect presented	audience reached
DIA	*Online Dissemination of CoeGSS flyer & homepage*		

ATOS	*Internal communication and publicity (intranet, internal bulletin, etc.).*		

Table 4 – News items, press releases and similar communication items

6 Outlook

In cooperation with T6.1 on community building, the tasks presented here will continue to work towards disseminating CoeGSS results and interacting with potential stakeholders, through events and training, so as to support the building of a sustainable Centre of Excellence for Global Systems Science. Next steps for the different tasks include the following:

- T6.2: the upcoming International Conference: Computing Power for Global Challenges, 24-25 October in Lucca, Italy, provides another opportunity of community building, dissemination, collaboration, and initiating new collaborations that only face-to-face events can present.
- T6.3: the production of an online training course shall provide a training opportunity for a wide audience, and shall open potential roads for further training initiatives, beyond the end of the project itself
- T6.4: monitoring how CoeGSS can contribute to standards, either with reference implementations or by providing input to the standards descriptions
- T6.5: presentations of CoeGSS content and results in conferences and workshops, for example in the GSS and HPC communities but also in events related to the pilot topics.

Overall, these tasks shall continue close cooperation within the project, with the activities of T6.1 on community building, with the pilots that drive the project, and with the work package concerned with the sustainability of the Centre of Excellence for Global Systems Science.

7 References

D2.2 – Final Sustainability Model

D4.1 – First Report on Pilot Requirements

D4.2 – Second Report on Pilot Requirements

D4.4 – First Status Report of the Pilots

D4.5 – Second Status Report of the Pilots

D6.3 – Second Annual Report on Community Building

D6.5 – First Annual Report on Training, Standardisation, Collaboration, Dissemination and Communication