# HiDALGO

## HPC and Big Data Technologies for Global Challenges


**Abstract**

Understanding major global challenges (GCs) as well as their underlying parameters is a vital issue in our modern world. The importance of assisted decision making by addressing global, multi-dimensional problems is more important than ever.

To predict the impact of global decisions with their dependencies, we need an accurate problem representation and a systemic analysis. To achieve this, HiDALGO enables highly accurate simulations, data analytics and data visualization, and also provides knowledge on how to integrate the various workflows as well as the corresponding data.
Our project aims to bring together the HPC, HPDA, and Global Systems Science (GSS) communities in order to address GCs and bridge the gap between traditional HPC and data-centric computation.



### Acknowledgements

- Runtime: December 1, 2018 - November 30, 2021
- Total budget: 7 991 500.00 EUR
- Consortium: 13 partners from 7 different countries
- Coordination: Atos Spain SA
- Technical management:
  High Performance Computing Center Stuttgart (HLRS)
- Researchgate:
  https://www.researchgate.net/project/HiDALGO-HPC-and-Big-Data-Technologies-for-Global-Systems



www hidalgo-project.eu

U.Project.HiDALGO  EU_HiDALGO  contact@hidalgo-project.eu


### Participating partners

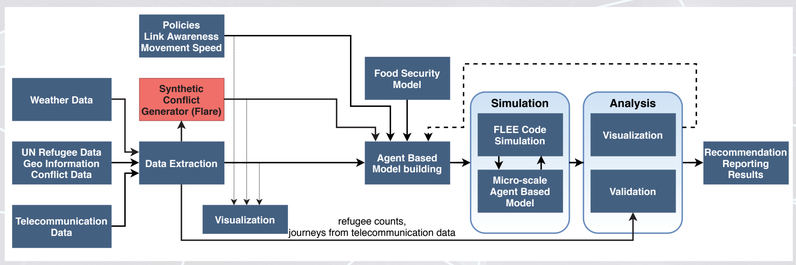
Atos · University of Salzburg · Brunel University London · PSNC · Széchenyi István University · Dialogik · Moonstar Communications · HLRS · ICCS · Magyar Közút · ARH · ECMWF · Know Center

### Motivation and main objectives

*Goals:*
- Benefit from the synergy between HPC, HPDA, AI, and GSS
- Connect & train GSS and HPC communities
- Provide a single entry point for decision makers, technical experts, and other relevant entities in the GCs ecosystem

*Project Targets:*
- Establish the baseline for HPC, HPDA and AI oriented computing in the domain of GCs
  - Advance state-of-the-art mechanisms for data analytics
  - Develop an AI-based integrated simulation workflow
- Focus on highly accurate models & significantly improved simulation results
  - Build-up coupled simulations for highly complex phenomena & advance coupling mechanisms
  - Integrate real-world data in static simulations
- Enhance and advertise a multi-domain portal for the GSS community

### Forced migration simulation

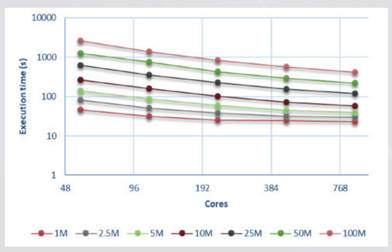



*Workflow for the migration use case*

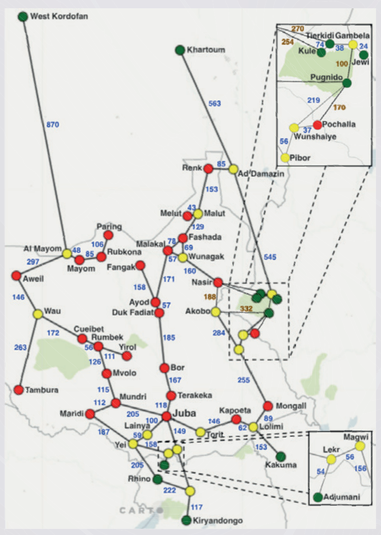
*Goals:*
- develop realistic models for simulating refugee streams
- complete data collection on refugee movements
- investigate the consequences of a nation closing its borders

*HPC in migration use case:*
- computationally expensive simulations: weather forecasting, detailed agent based models, etc.
- ensemble runs
  - analysis of many scenarios
  - VVUQ due to probabilistic nature of ABMS and uncertainties in input data

*HPDA in migration use case:*
- validate models and measure uncertainties in simulation results (statistical inference)
- generate high resolution synthetic populations from weak census data and microdata (probabilistic graphical models, IPF, etc.)
- process geo-spatial information
  - extract list of relevant refugee locations
  - compute likely walking and driving routes between these locations

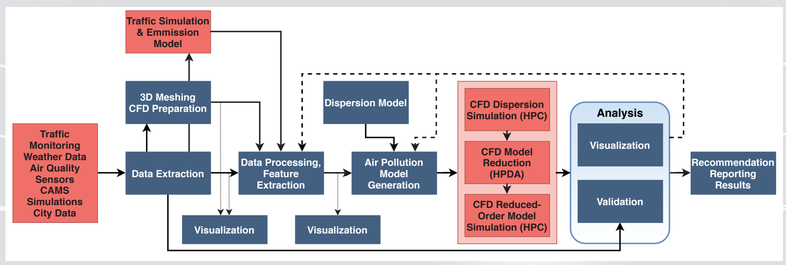



*Scalability of Flee ABMS code for simulating human migration*




*Geographic network model for South Sudan conflict*
*D. Suleimenova & D. Groen (2020): JASSS 23 (1), 2. DOI: 10.18564/jasss.4193*

### Urban air pollution simulation

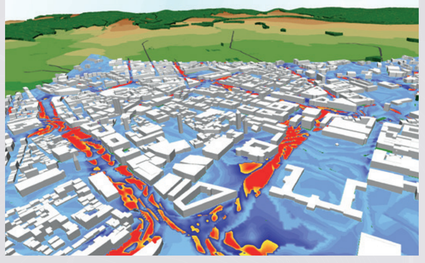



*Workflow for the UAP use case*

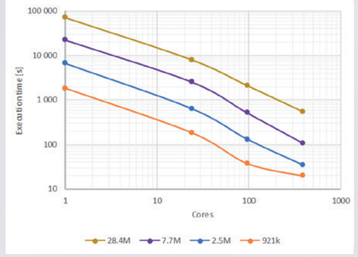
*Goals:*
- simulate pollution in cities based on real-world sensor data
- couple models for traffic, air flow, and weather simulation
- provide stakeholders with evidence based decision models to leverage green growth

*HPC in UAP use case:*
- computationally expensive simulations: weather forecasting, detailed CFD models of the pollution spread in the city, etc.
- ensemble runs
  - analysis of many scenarios
  - VVUQ due to uncertainties in input data

*HPDA in UAP use case:*
- validate models and measure uncertainties in simulation results (statistical inference)
- reduce models for faster simulation (PCA, SVD)
- impute the missing data from sensors (Bayesian networks, etc.)

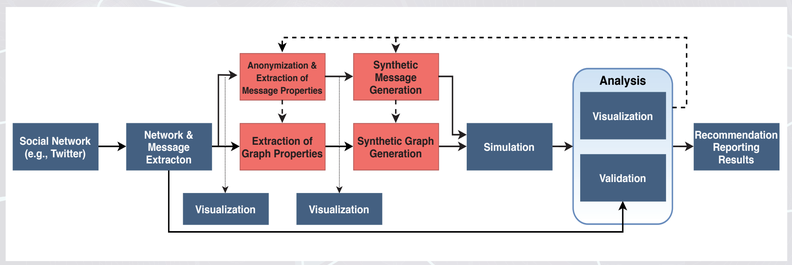


*UAP simulation results for the city of Stuttgart visualized with Covise*




*Scalability of the UAP simulation implemented with OpenFOAM framework*

### Social network analysis




*Workflow for the SNA use case*

*Goals:*
- analyze the structure of social networks
- analyze the stochastic behavior of the message spreading
- simulate the spread of messages among users
- study the spread of malicious messages and develop countermeasures

*HPC in SNA use case:*
- computationally expensive network analytics algorithms
  - estimating the spectra of large scale networks for verification of social network models
  - computationally expensive methods for clustering of large scale networks
- computationally expensive simulations

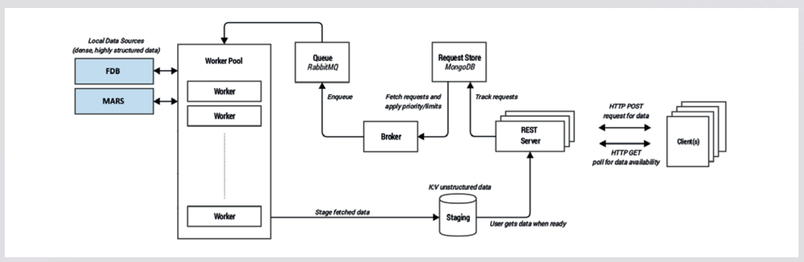
*HPDA in SNA use case*:
- extract features from Twitter datasets
  - more than 20 features were chosen from the analysis
- model the retweet probabilities (logistic regression, tree-based method)

*All scalability results are provided for Hazel Hen supercomputer, a Cray XC40 system installed at HLRS.*

*Scalability of pilot applications*

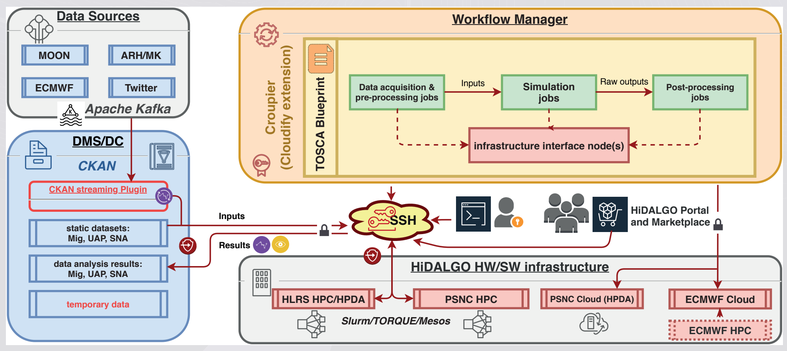
| Metric | Migration | UAP | SNA |
|---|---|---|---|
| elapsed time in hours (production run) | 1 | 72 | 3.3 |
| max cores per run (in production) | 448 | 384 | 7200 |
| # of runs (in production) | 500 | 40 | 90 |
| max cores per run (scalability test) | 896 | 432 | 7200 |
| max speedup per run | 300 | 200 | - |

### Generalized workflow and infrastructure

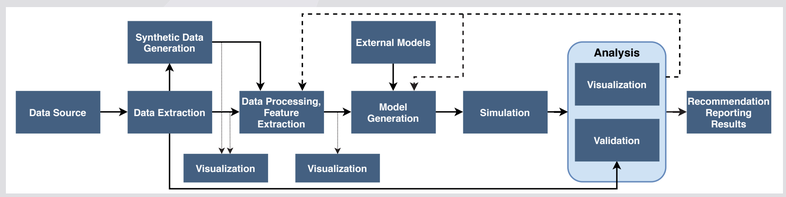



*Preliminary design for ECMWF REST API*

- Cloudify with Croupier plugin for orchestrating & monitoring
- CKAN as a data management system
- Apache Kafka for integration of streaming data
- custom REST API implementation for coupling of data sources and simulations across data centres (see ECMWF design for coupling with weather and climate data)




*High level system architecture and workflow*




*Generalized workflow for GSS applications*

### Further work and challenges to address

- improve scalability in the simulation tools and HPDA components
  - identify bottlenecks via benchmarking
  - analyze & improve I/O operations
  - analyze potential benefit from using accelerators (GPUs)
  - evaluate HPC/HPDA tools for GSS domain on innovative architectures
- enable coupled simulations
  - complete REST-based weak coupling
  - implement strong coupling mechanisms (via messages passing)
  - introduce coupling with streaming data
    - evaluate tools for processing data coming from IoT
    - enable data streams in HPC
  - improve mechanisms for moving large datasets


The HiDALGO project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824115