Analysis pipeline for hydrological simulation in the PYRAMID project

Robin Wardle, Mark Turner

Newcastle Data, Newcastle University

Systems Research Challenges Workshop, 30th Nov - 2nd Dec 2021

From Newcastle. For the world.

PYRAMID Demonstrator Platform

PYRAMID: Platform for dYnamic, hyper-resolution, near-real time flood Risk AssessMent Integrating repurposed and novel Data sources

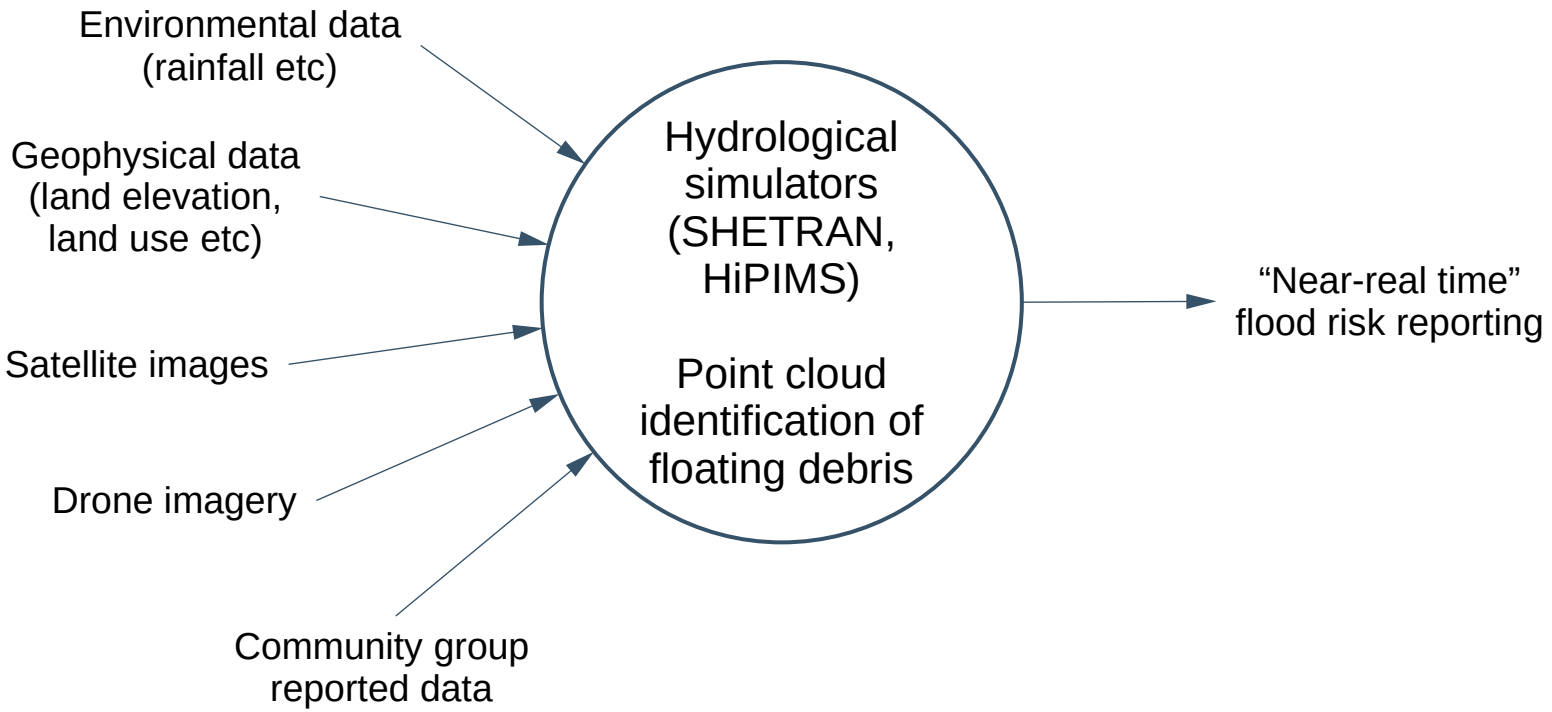
Collaborators

- Newcastle University Civil Engineering, Water Research Group
- Newcastle University Urban Observatory
- Loughborough University Civil Engineering
- Newcastle Data
- External stakeholders: Newcastle City Council, National Rail, Highways Agency, Residents and Community groups

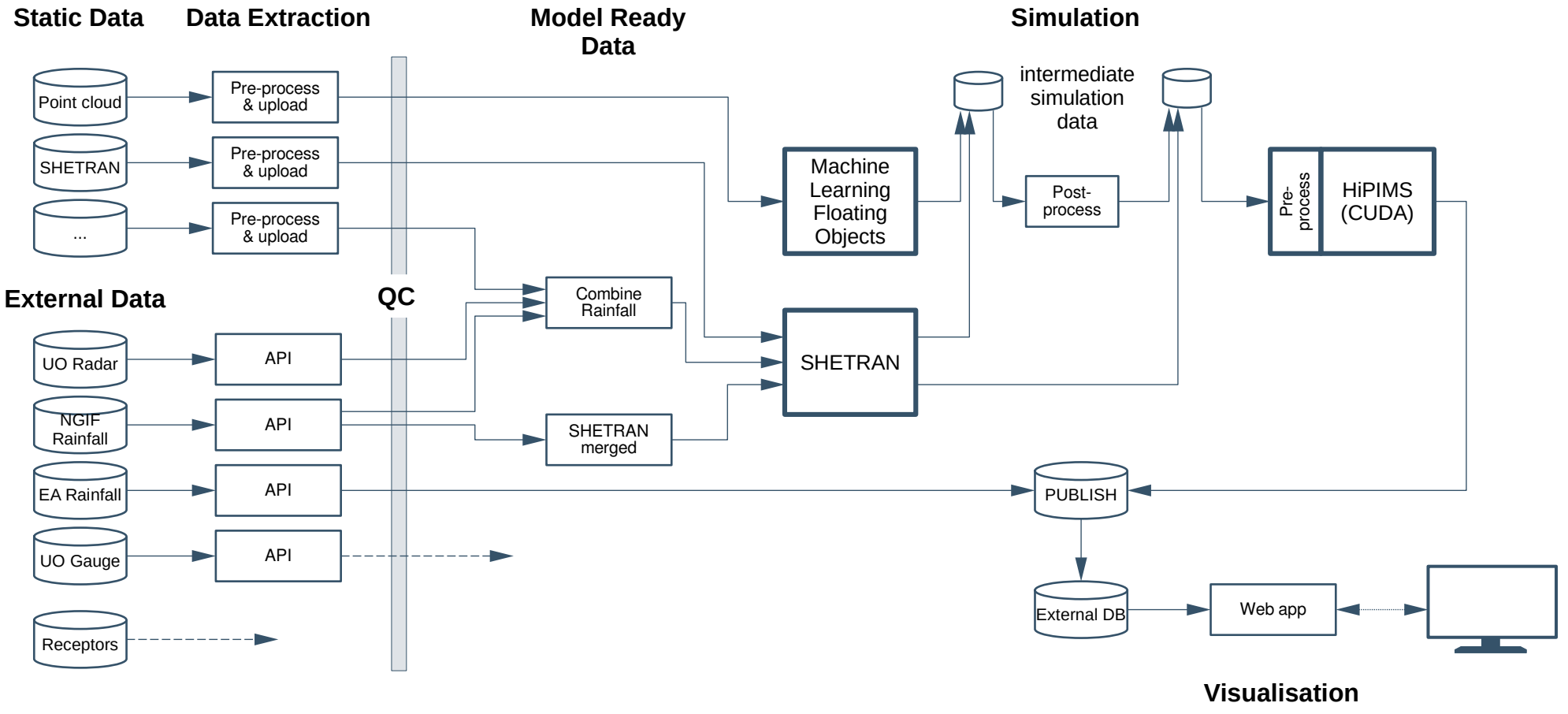
Problem

Photo sources: BBC News (Steven Lomas), ChronicleLive, Railfreight, Urban Foresight

Research Project



Analysis pipeline for hydrological simulation in the PYRAMID project



Constraints and Requirements

Project is fairly advanced along its schedule

Limited staff resource

This kind of problem occurs regularly in a research environment

So …

Looking for a pragmatic solution

Re-using expertise and existing technology where possible

Solution

DAFNI - Data & Analytics Facility for National Infrastructure

Positives

1) Containerised applications using Docker
2) Existing mature(-ish) platform built around Kubernetes
3) Huge amounts of dataset storage
4) Free for academic use (currently)
5) Reasonable compute facility (200 nodes + 10 GPU nodes)
6) Good support (direct, Slack)
7) Builds on existing expertise within the university
8) Don’t need to build our own framework from scratch
9) Project groups, data and models search
10) GitHub integration

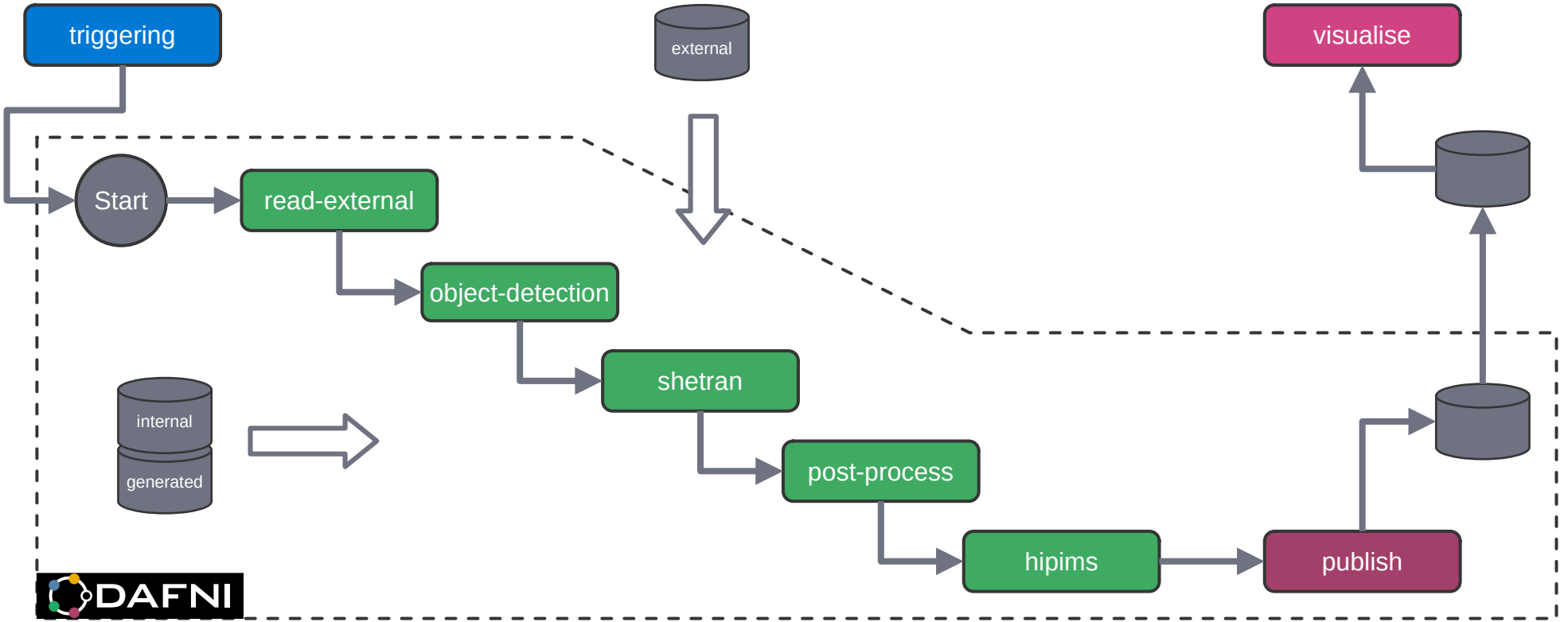
Limitations

1) Poorly documented API
2) Triggering tasks can’t be done within the platform
3) Visualisation will be an external component
4) WIP
5) Workflow interface is weak
6) Laborious UI
7) Downtime and outages
8) Some features only available on request
9) Internal data transfer limitations (~30GB)
10) No job scheduling, coordination or parallel job execution

DAFNI

https://dafni.ac.uk

Workflow Design in DAFNI



Conclusions

DAFNI will provide the backbone of the PYRAMID demonstrator system – container management, workflow orchestration, static data storage

Additional – dynamic – data will come from external APIs

Microsoft Azure Services will be used to deploy additional software which will coordinate the execution of the DAFNI pipeline and provide a visualisation service

Use the DAFNI API to construct the workflow itself (infrastructure as code)

Version 2?

Thanks

Contact

Robin Wardle, Research Software Engineer <robin.wardle@newcastle.ac.uk>

Mark Turner, Head of Research Software Engineering <mark.turner@newcastle.ac.uk>

Elizabeth Lewis, Deputy Director of Research & Innovation, <elizabeth.lewis2@ncl.ac.uk>