Analysis pipeline for hydrological simulation in the PYRAMID project

Robin Wardle

Research Software Engineer

Newcastle Data, Newcastle University

DAFNI Conference 2022 Environmental Impacts, 5th Jul 2022, Manchester UK

Newcastle University

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

Environmental Problem

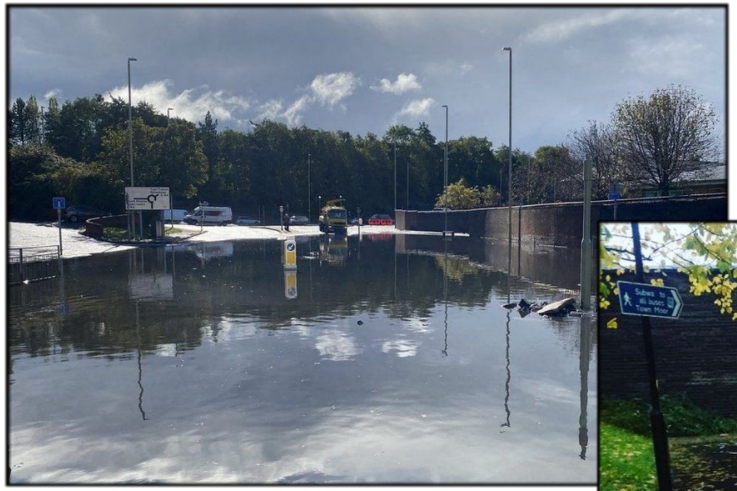
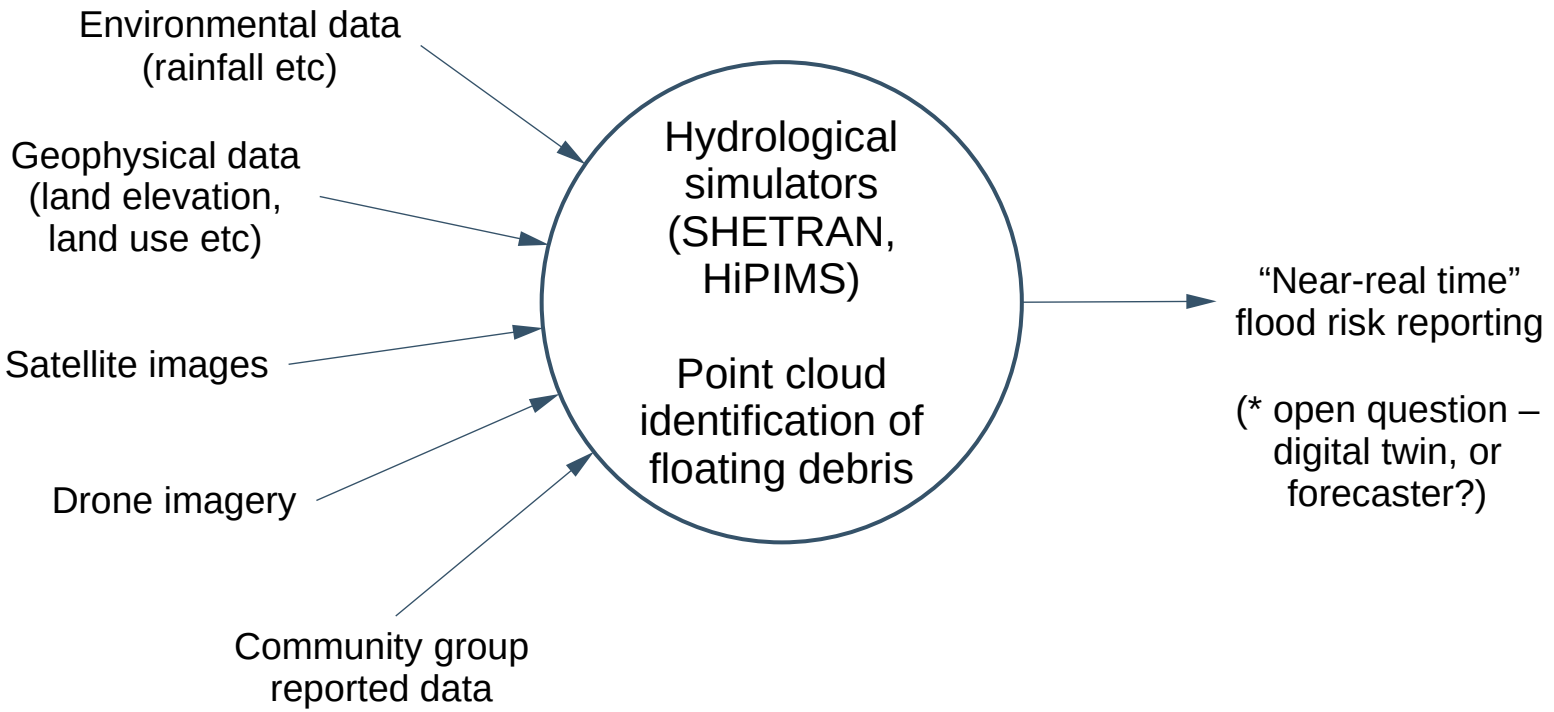
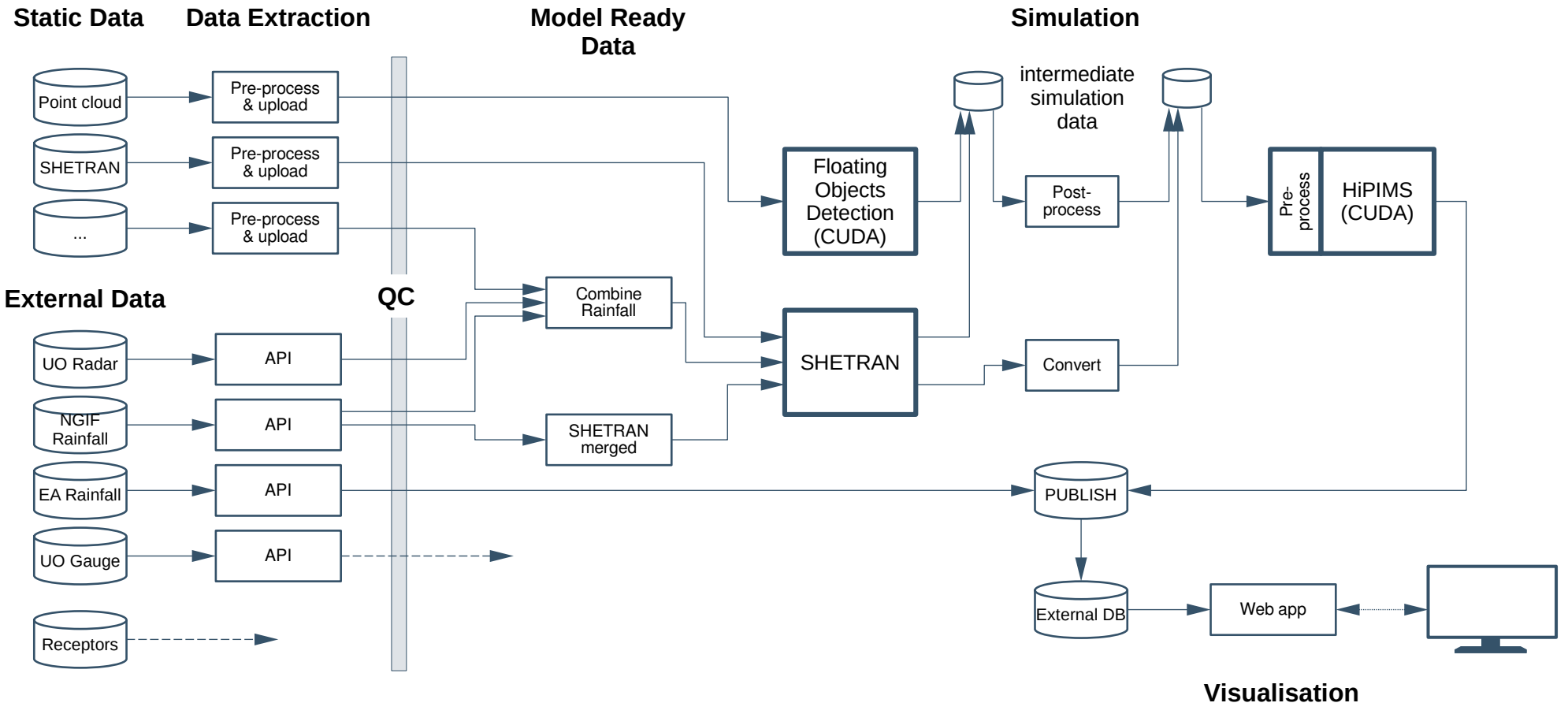


Photo sources: BBC News (Steven Lomas), ChronicleLive, Railfreight, Urban Foresight

Research Project



Analysis pipeline for hydrological simulation in the PYRAMID project



Constraints and Requirements

Project was / 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 and development roadmap
3) Huge amounts of dataset storage
4) Free for academic use (currently)
5) Reasonable compute facility (200 nodes + 10 GPU nodes)
6) Great 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 and automation

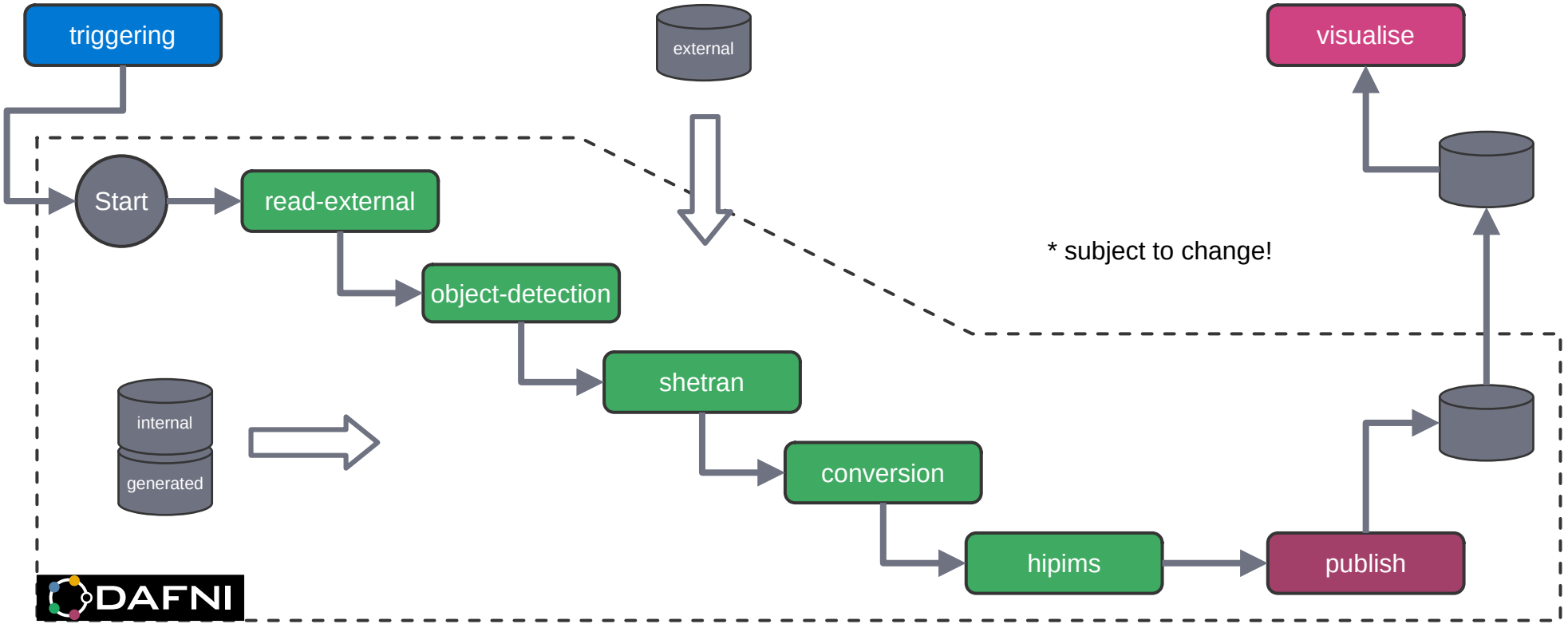
Limitations

1) Poorly documented API / unreleased CLI
2) Job scheduling can’t be done within the platform
3) Visualisation is an external component
4) WIP (currently v3.1.0)
5) Workflow interface is limited
6) Laborious UI
7) Downtime and outages
8) Some features only available on request
9) Internal and external data transfer limitations (~30GB)
10) No parallel job execution

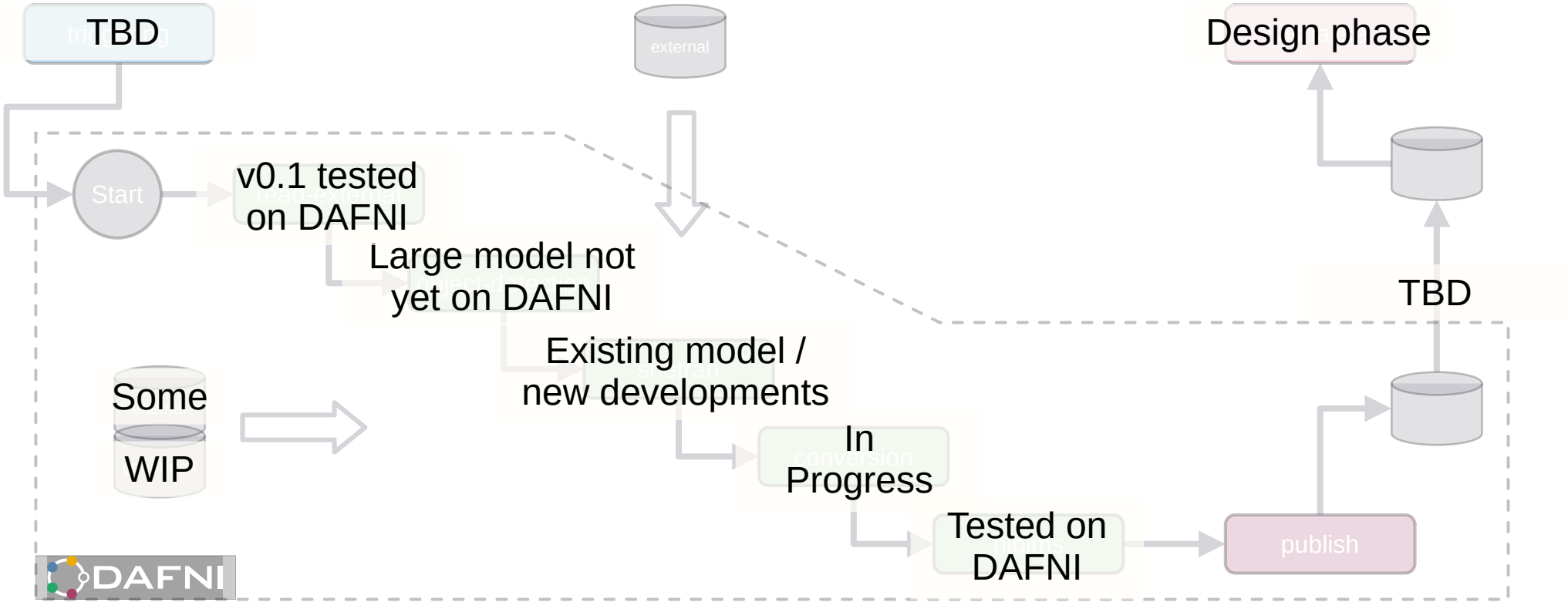
DAFNI

https://dafni.ac.uk

Workflow Design in DAFNI



Progress


TBD
external
Design phase
Start
v0.1 tested on DAFNI
Large model not yet on DAFNI
TBD
Existing model / new developments
Some
WIP
In Progress
Tested on DAFNI
publish
DAFNI

Challenges

HiPIMS and the Object Detection models use GPUs (CUDA)

- Docker image creation is very complicated for these models
- Images are 7GB and 14GB respectively and break DAFNI ingestion

External job triggering and data collection not yet designed

Docker image creation and DAFNI asset version management still confusing for non-specialists

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

DAFNI v3.1.0 features to be explored

Multi-GPU use

Error reporting and logging from Python scripts to Sentry (https://sentry.io/)

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

Thanks

PYRAMID
https://gtr.ukri.org/projects?ref=NE/V00378X/1

SHETRAN
https://research.ncl.ac.uk/shetran/

HiPIMS
https://www.hemlab.org/models/

Contact

Robin Wardle, Research Software Engineer <robin.wardle@newcastle.ac.uk>
Elizabeth Lewis, Deputy Director of Research & Innovation <elizabeth.lewis2@ncl.ac.uk>