

# **A proposal for rapid model development system using existing open data services**

*Realising the full benefit of open data policy*

Janne Ropponen & Janne Juntunen,  
Finnish Environment Institute,  
Modelling seminar, 1.4.2014

## Contents

- Overview of existing services for data and modelling
- A new way to build models using open data
- How to use the system – an example

## Some remarks

- The ideas presented here stem from the experiences and hardships during everyday work with 3-D lake models and 1-D river models
- The viewpoint of this presentation: *How can I as an environmental modeller make my work more productive and beneficial to more people?*
- The modelling framework presented here does not yet exist. All comments, ideas and co-operation are warmly welcomed to make this a reality.
- **THINK BIG – there are no limits!**

## Some environmental open data sources in Finland

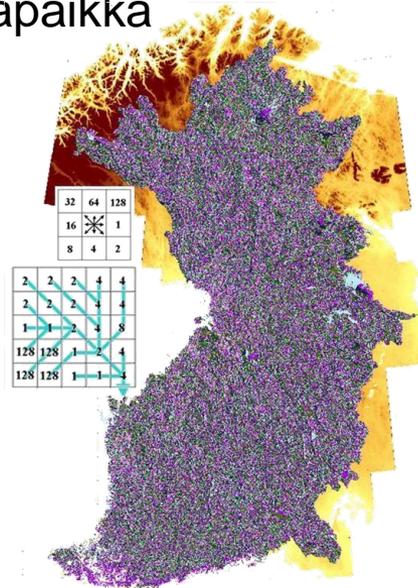
- With the EU INSPIRE-directive, the availability of georeferenced open data from public organisations has increased significantly
  - **SYKE** Hydrological and water quality data, bathymetry data, waterways, ...
  - **FMI** Marine, weather and air quality data & forecasts
  - **NLS** Topographic datasets
  - **GTK** Geology data etc
  - **Statistics Finland** Statfin, Eurostat databases
  - ...and many more
- 
- <http://data.suomi.fi/>
  - <http://beta.avoindata.fi/>
  - <http://www.ymparisto.fi/oiva>
  - <http://fmi.fi/avoin-data>
  - <http://www.maanmittauslaitos.fi/en/opendata>
  - <http://www.openstreetmap.org/>
  - <http://stat.fi/tup/>
  - <http://hakku.gtk.fi/>



# Open data services and databases

**Tools & (upcoming) projects – everyday services which combine data and user interface**

- JärviWiki
- Paikkatietoikkuna
- Kansalaisen karttapaikka
- LifeData
- EnviBase
- Google Maps
- OpenStreetMap
- Oiva/Hertta
- LAPIO
- Vesinetti
- VALUE
- Lynet/Paikkatieto
- ...



FMI DATA

# Free, publicly available model building systems

-  **Opasnet** <http://opasnet.org>
  - Maintainer: THL
  - Model building & running with R inside a Wiki-environment
  - Online
-  **ESMF** <http://www.earthsystemmodeling.org/>
  - Supporters: NASA, DoD, NSF, NOAA
  - Framework for building coupled modeling systems
  - Open source
  - Offline
-  <http://openearth.eu/>
  - Maintainers: Deltares, Delft TU, etc.
  - Open source project for data, models and tools
  - Partly online, partly offline

# Open modelling components



- **OpenMI** <http://www.openmi.org/>
  - **Open Modelling Interface**
  - Standard way for exchanging information between dynamic models



- **OpenDA** <http://www.opendata.org/>
  - Open data assimilation and calibration tools for models
  - Numerous methods already implemented and ready-to-use

## Deficiencies of current model-building systems

- Few really take advantage of open data
- No combined service to access, convert and apply data to models
- Existing systems are often confined to single purpose and aren't free nor open enough

## Solution: An open, online modelling system

- The goal: *To eliminate all the boring, repetitive data gathering, processing and code implementation work when building models and concentrate resources on improving the model*
- Majority, if not all, of the components to this end are already available – all that needs to be done is combine them together

# (Environmental) Modelling system

## Traditionally

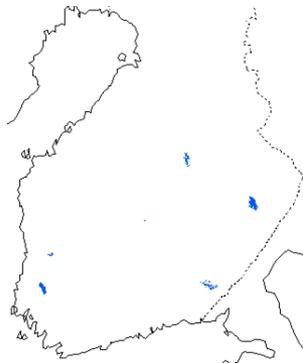
1. Choose target site
2. Build model simulation piece-by-piece
  - Get data, process data, build code, choose parameters, ...
3. Refine, validate model
4. Report results

## In the future

1. Choose target site
  - Model built automatically by the system
2. Refine, **develop**, validate model
3. Report results
4. (Update modelling system)

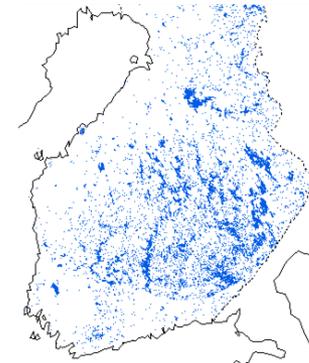
## Results

1. A few detailed, local models
2. All other sites remain unmodeled
3. Reports/results often hard to find



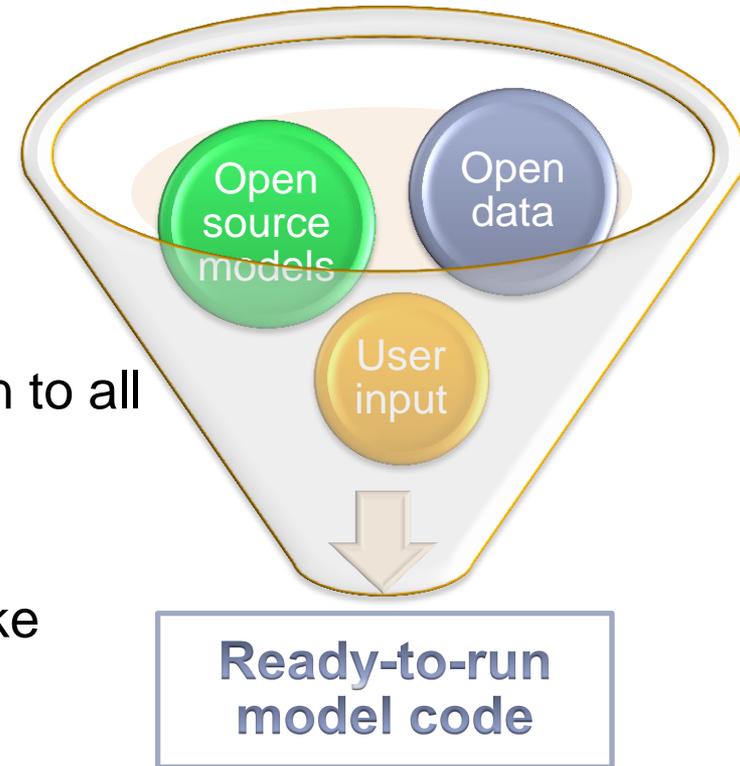
## Results

1. A first-guess model for all sites
2. Multiple detailed local models
3. Models and results shared with all



## Modelling system – the basics

- **Open data driven** – uses existing data sources
- Applies open data to open source model codes **on-demand**
- Not limited to just experts – open to all
- Open access, open source and open interfaces – users can make the system better
- A **tool** for (environmental) model building



# Modelling system – the basics

- The system **automates** the tedious **tasks** of
  - Data gathering and processing
  - Basic model building and chaining
- The system **doesn't** (though added components can!)
  - Store data – rather the data is fetched from existing sources
  - Run models
  - Check, fine-tune and validate models
- The system **can be**
  - Developed by everyone (so it doesn't die when funding does)
  - Updated and upgraded (new models, connections and data)

## Some arguments why this should be done

- Finland's good quality open data systems provide for a great testbed for the system
- Saved work hours from basic model building tasks can be used to actually improve the models and code new features
- Easy to use system will immediately benefit researchers, consultants and the public by lowering the bar for modelling
- Expertise of professionals are directly shared with as large an audience as possible
- Open platform encourages people to share their models and data
- Because modelling tools are made freely available authorities can actually start to require model based estimations as standard when doing e.g. environmental impact assessments

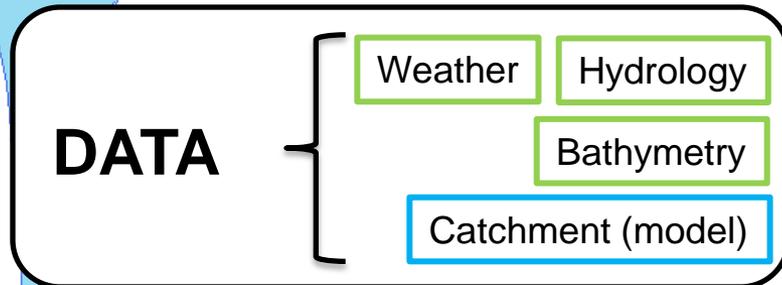
## Other ideas on how to use the framework

- Cloud computing: Although models are not run within the framework, the framework can send the models to the cloud for runs
- Commercial services: Consultants can sell their data through the system – ready to use in modelling
- Your idea?

## Modeling system in use – an example

- **Problem:** What is the economic impact of health problems caused by contaminants in wastewater making their way into drinking water?
- **Solution** using the modelling framework GUI:
  1. Define models to use on relevant areas: E.g. lake model → river model → water purification model → health model → economic model
  2. Define wastewater plant, drinking water plant to use
  3. Define time period of interest
  4. Define scenarios to model
  5. The system automatically retrieves and processes all the data relevant to the various models and builds the model chain
  6. Running the chain of models gives the cost to the economy in each of the scenarios as the end result

# Example: Chaining models



Contaminant input

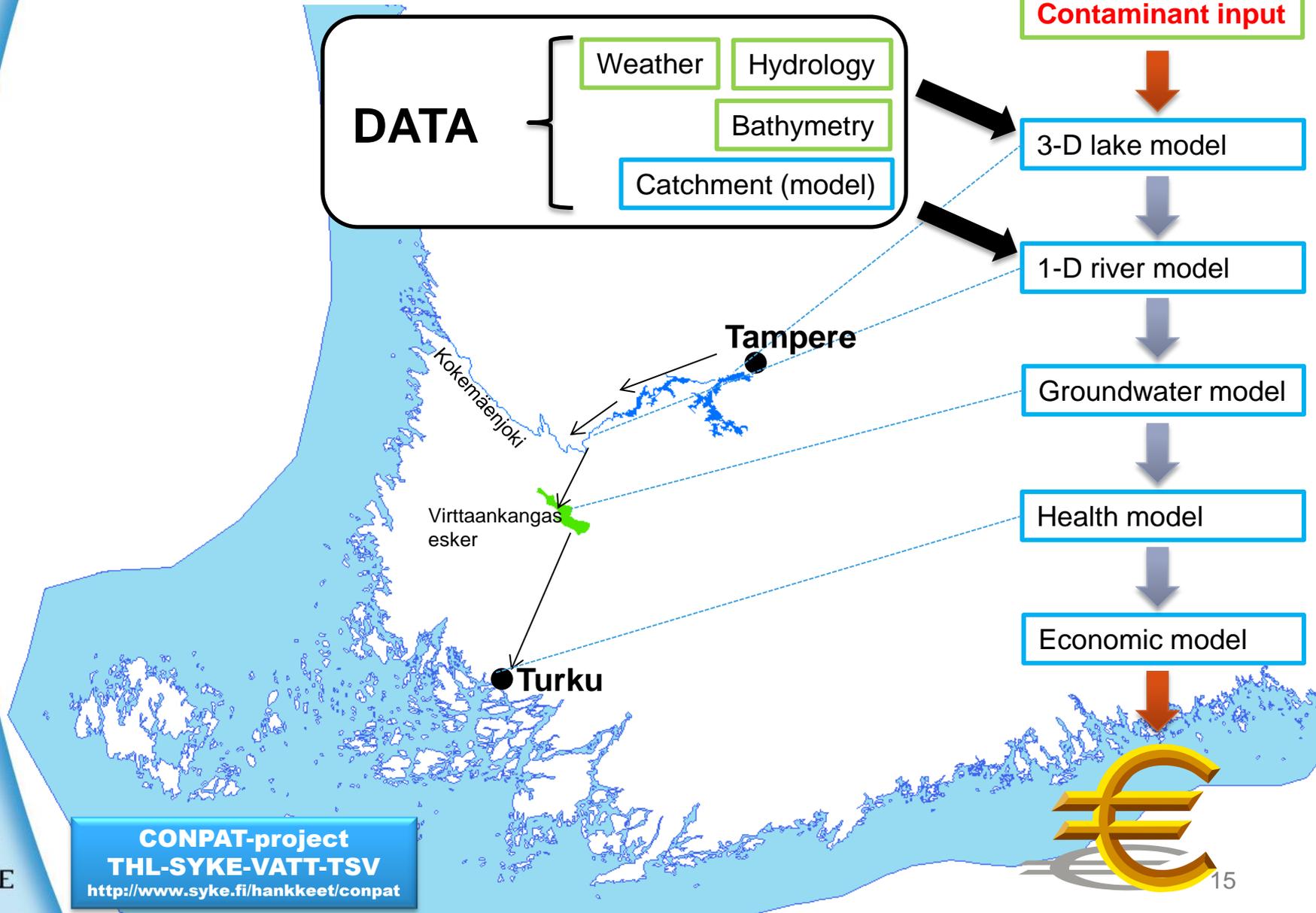
3-D lake model

1-D river model

Groundwater model

Health model

Economic model



## Summary

- A unified, easy-to-use tool is needed for rapid environmental modelling needs
- Solution is to develop an open interface, open source map-based system that automatically builds first-guess models using open data sources
- The system is designed to be built incrementally
- Most of the components needed to build such a system are already available – and are both open and free

## Thank you – kiitos!

Questions, remarks, similar projects,  
co-operation, ..., ?

Contact:

- [janne.ropponen@ymparisto.fi](mailto:janne.ropponen@ymparisto.fi)
- [janne.juntunen@ymparisto.fi](mailto:janne.juntunen@ymparisto.fi)

