Loughborough University

Steven K Firth, Gareth Cole, Tom Kane, Farid Fouchal & Tarek M Hassan

An open data science approach for building performance studies using refitXML and Jupyter Notebooks

Presentation to the eSim 2018 conference, Montreal, 9-10 May 2018

How can I write an Open Science building simulation journal paper?

Contents

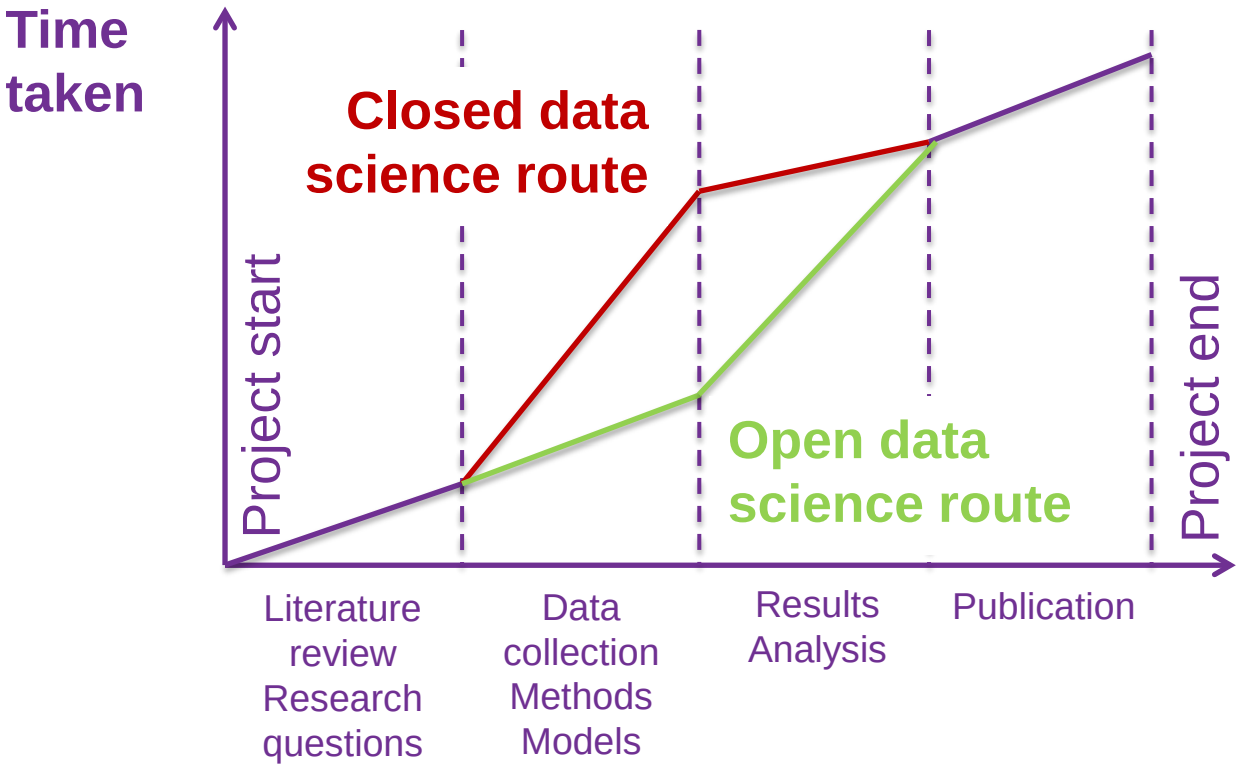
- Open Science
- Workflow for building performance studies
- Data structures vs. analysis complexity
- Open Data – a case study using refitXML and figshare
- Open Methodology – a case study using Python, Jupyter Notebooks and GitHub
- Conclusions & next steps

What is Open Science?

1. Open Access: publishing the results of academic research as freely available on the public internet
2. Open Data: publishing the datasets collected in the research process, without restricting their use
3. Open Source: making software developed in the research available under an open license
4. Open Methodology: sharing the methodology of a study, and the tools used for data collection and analysis

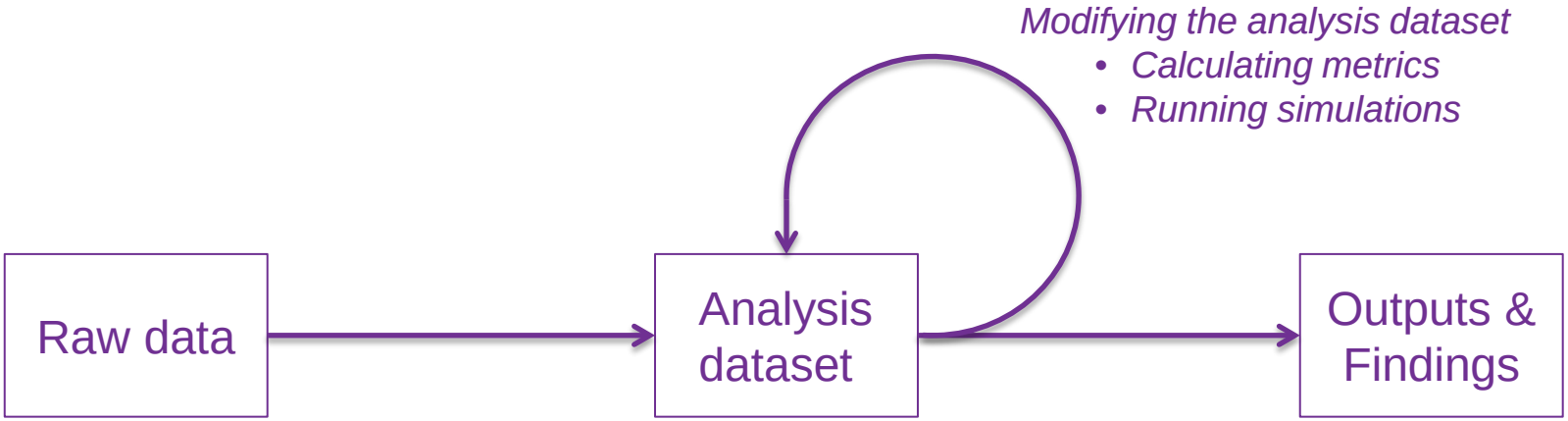
Kraker P., Leony D., Reinhardt W. and Beham G. 2011.
The case for an open science in technology enhanced learning,

Benefits of Open Science



Steps of a typical building performance research project

Workflow for building performance studies



Building surveys
Architect drawings
Questionnaires
gbXML files
Sensor data
Simulation model files
Weather data
cityGML files

1. Simulation input data
   - *geometry, constructions, internal heat gains*
   - *airflow, HVAC*
2. Descriptive data
   - *construction age, built form type*
   - *occupant age, incomes, professions*
3. Time series data
   - *Sensor measurements*
   - *Simulation results*

Tables
Plots
Images
Data files

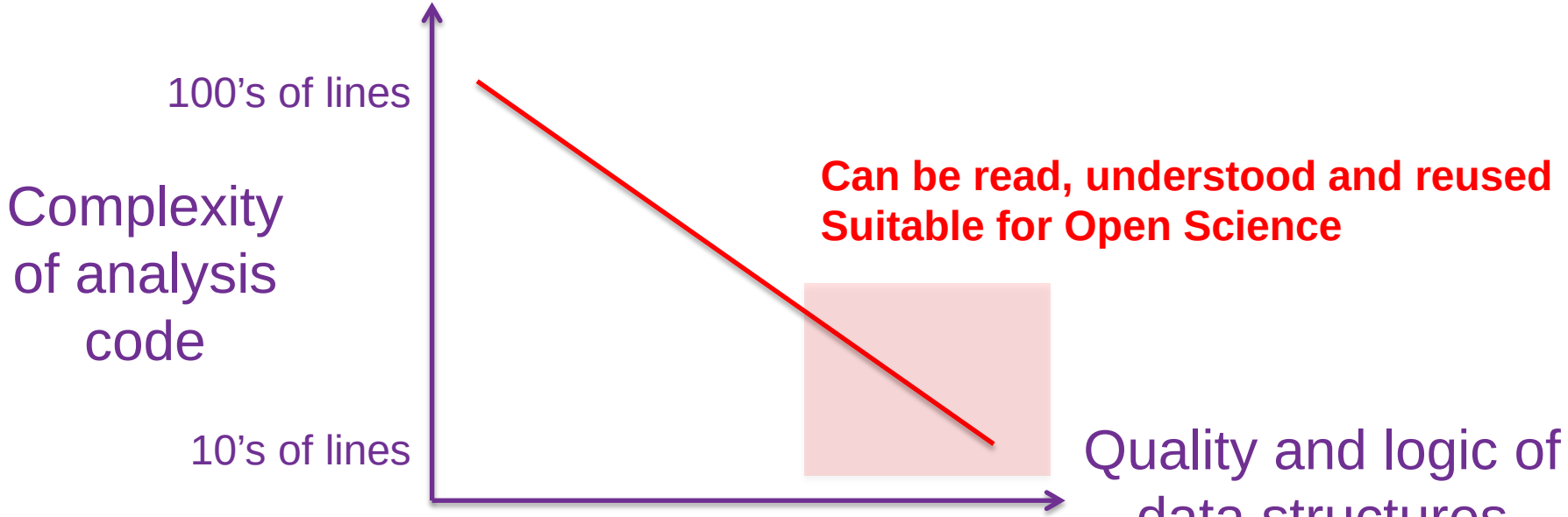
Data structures

- “I will, in fact, claim that the difference between a bad programmer and a good one is whether he considers his code or his data structures more important. Bad programmers worry about the code. Good programmers worry about data structures and their relationships.”

Linus Torvalds

(as quoted in ‘The Self-Taught Programmer’ by Cory Altoff)

Data structure quality vs. analysis complexity



Open Data – a case study

- Loughborough University data repository
- REFIT Smart Home Dataset
- refitXML file – for building survey data
- csv file – for sensor data

Loughborough University

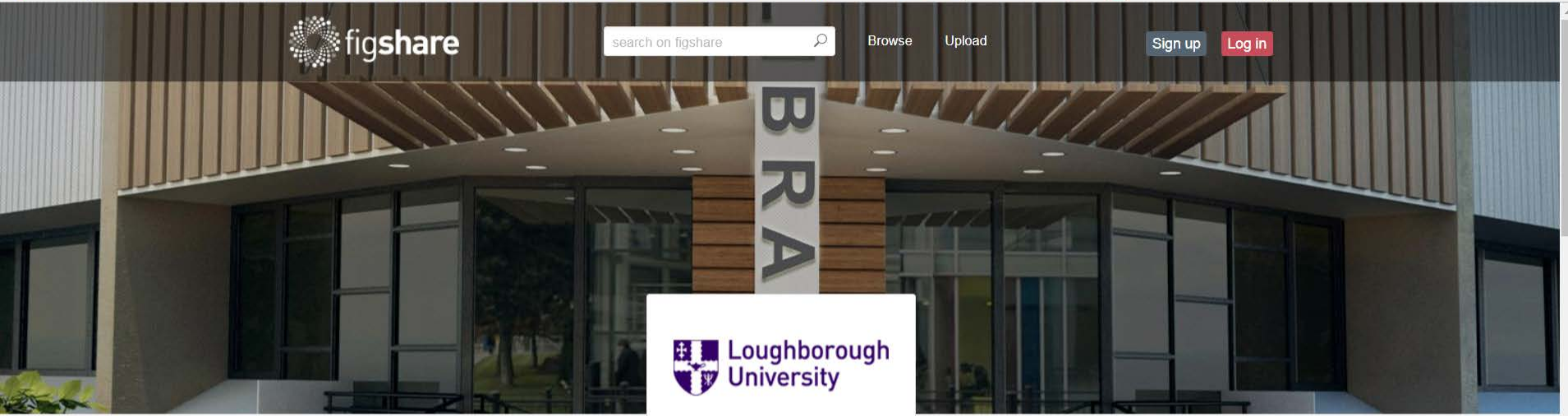
Secure | https://lboro.figshare.com

Apps | Current Students an | Meeting Room Book | CoTutor | P | Figshare | GitHub | Python | REFIT | REFIT ‹ Log In | Firth, Steven | Civil a | ISO 52016-1:2017 - | Unit 6: Simple Linea

figshare

search on figshare

Browse Upload Sign up Log in



Loughborough University

Discover research from Loughborough University

NEW POPULAR CATEGORIES SEARCH

65682 views | **14020** downloads | more stats...

CONFIDENTIAL FILE

11 months

FILE UNDER EMBARGO

FILESET

ReadMe.txt (6.83 kB)

RefitXMLSchema.xsd (90.84 kB)

REFIT_BUILDIN... .xml (1.67 MB)

REFIT_TIME_SE... .zip (94 MB)

Cite | Download all (95.77 MB) | Share | Embed | + Collect (you need to log in first) | 4 files

REFIT Smart Home dataset

20.06.2017, 11:53 by Steven Firth, Tom Kane, Vanda Dimitriou, Tarek Hassan, Farid Fouchal, Michael Coleman, Lynda Webb

This dataset is maintained by Steven Firth (s.k.firth@lboro.ac.uk), Building Energy Research Group (BERG), School of Civil and Building Engineering, Loughborough University.

The REFIT project (www.refitsmarthomes.org) carried out a study from 2013 to 2015 in which 20 UK homes were upgraded to Smart Homes through the installation of devices including Smart Meters, programmable thermostats, programmable radiator valves, motion sensors, door sensors and window sensors.

Data was collected using building surveys, sensor placements and household interviews.

The REFIT Smart Home dataset is one of the datasets made publically available by the project. This dataset includes:

- Building survey data for the 20 homes.

refitXML

```
<Building id="Building01" startDateTime="2013-10-01T00:00:00Z" occupancyType="Single
    family dwelling" builtFormType="Detached house or bungalow" orientation="327"
    wallTypeMainBuilding="Masonry-Boxwall-Cavity" wallAgeBandMainBuilding="1975 - 1980"
    cavityWallInsulationPresent="Yes" windowType="Double glazed - UPVC" loftType="Fully
    boarded" loftInsulationType="Mineral wool/fibre glass" loftInsulationThickness="300mm">
    ...
    <Space id="Space1" startDateTime="2013-10-01T00:00:00Z" conditionType="Heated"
        area="6.25" volume="14.375" storeyLevel="0" roomType="Study">
        <Sensor id="Sensor41" startDateTime="2013-10-02T05:00:00Z" endDateTime="2013-
            12-03T15:15:00Z" manufacturer="Onset" model="Hobo pendant">
            <TimeSeriesVariable id="TimeSeriesVariable41" startDateTime="2013-10-
                02T05:00:00Z" endDateTime="2013-12-03T15:15:00Z" variableType="Air
                temperature" units="C" intervalType="FixedInterval" intervalUnit="Minute"
                intervalLength="15" hasMissingData="No" repeatsOmitted="No"
                hasDuplicateTimestamps="No"/>
        </Sensor>
        ...
    </Space>
    ...
</Building>
```

refit csv

TimeSeriesVariable/@id	dateTime	Data
TimeSeriesVariable1	2013-10-02T05:00:00Z	17.772
TimeSeriesVariable1	2013-10-02T05:30:00Z	18.081
TimeSeriesVariable1	2013-10-02T06:00:00Z	18.176
TimeSeriesVariable1	2013-10-02T06:30:00Z	18.176
TimeSeriesVariable1	2013-10-02T07:00:00Z	18.105
TimeSeriesVariable1	2013-10-02T07:30:00Z	18.01
TimeSeriesVariable1	2013-10-02T08:00:00Z	17.891
TimeSeriesVariable1	2013-10-02T08:30:00Z	17.772
TimeSeriesVariable1	2013-10-02T09:00:00Z	17.701
TimeSeriesVariable1	2013-10-02T09:30:00Z	17.677
...	...	...

Open Methodology – a case study

- GitHub
- Repository for this conference paper
- Python and Jupyter notebooks

Steven Firth

stevenkfirth

Reader in Building Performance Modelling, School of Archtiecture Building and Civil Engineering

Edit bio

Loughborough University
Loughborough, UK
s.k.firth@lboro.ac.uk
http://www.lboro.ac.uk/depart...

Organizations

Popular repositories

Customize your pinned repositories

energy_savings_using_regression
This repository contains the analysis for the journal paper 'Energy savings using regression'
Jupyter Notebook

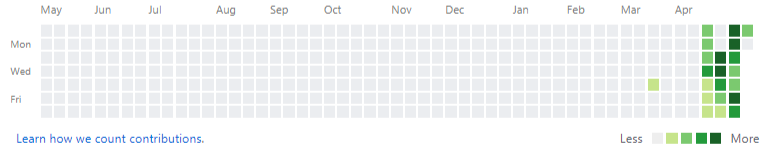
uSim_2018
Analysis for the uSim 2018 conference - Urban Energy Simulation - Glasgow, 29-30 Nov 2018
Jupyter Notebook

eSim_2018
Analysis for the eSim 2018 conference in Montreal, Canada, on 9-10 May 2018
Jupyter Notebook

DATALEB_2018
Analysis for the DATALEB 2018 conference in Murcia, Spain, on 28-29 June 2018

102 contributions in the last year

Contribution settings



Contribution activity

Jump to | 2018

May 2018

Created 19 commits in 4 repositories

openbuilding/openbuilding 9 commits
stevenkfirth/energy_savings_using_regression 6 commits
stevenkfirth/uSim_2018 2 commits

stevenkfirth / eSim_2018

Watch 0 | Star 0 | Fork 0

Code | Issues 0 | Pull requests 0 | Projects 0 | Wiki | Insights | Settings

Analysis for the eSim 2018 conference in Montreal, Canada, on 9-10 May 2018

Edit

Add topics

10 commits | 1 branch | 0 releases | 1 contributor | MIT

Branch: master | New pull request | Create new file | Upload files | Find file | Clone or download

stevenkfirth Update README.md — Latest commit 7c3ebd9 7 days ago

analysis	Create README.md	7 days ago
.gitignore	Initial commit	7 days ago
LICENSE	Initial commit	7 days ago
README.md	Update README.md	7 days ago

README.md

eSim_2018

Analysis for the eSim 2018 conference in Montreal, Canada, on 9-10 May 2018

This GitHub repository contains all the code used in the paper:

> *'An open data science approach for building performance studies using refitXML and Jupyter Notebooks'* by Steven K Firth, Gareth Cole, Tom Kane, Farid Fouchal and Tarek M Hassan, submitted to the eSim_2018 conference.

How to read this repository

- 'analysis': this folder contains the code used for the analysis in the paper. Look in this folder to see how the analysis is done

stevenkfirth / eSim_2018

Watch 0 Star 0 Fork 0

Code Issues 0 Pull requests 0 Projects 0 Wiki Insights Settings

Branch: master eSim_2018 / analysis / Create new file Upload files Find file History

stevenkfirth Create README.md Latest commit 75112ca 7 days ago

..		
README.md	Create README.md	7 days ago
summary_stats.ipynb	Added analysis	7 days ago

README.md

eSim_2018/analysis/

This folder contains the 'summary_stats.ipynb' Jupyter notebook used for the analysis. The code can be seen by opening this notebook.

To run this notebook:

1. download the notebook from GitHub (or clone this repository)
2. download the REFIT data from https://doi.org/10.17028/rd.lboro.2070091.v1
3. place the 'REFIT_BUILDING_SURVEY.xml' file and the 'REFIT_TIME_SERIES_VALUES.csv' in the same folder as the Jupyter notebook
4. run the notebook

Jupyter notebook

summary_stats

This notebook shows how to calculate summary statistics of the information stored in the REFIT Smart Home Dataset.

In [35]:
```
from lxml import etree; from collections import Counter; import pandas as pd  # imports external libraries and packages
xml=r'REFIT_BUILDING_SURVEY.xml'  # the name of the refitXML file
csv=r'REFIT_TIME_SERIES_VALUES.csv'  # the name of the refit csv file
NS={'a':'http://www.refitsmarthomes.org'}  # the xml namespace of the refitXML file
tree=etree.parse(xml)  # parses the refitXML file into an lxml etree variable
elements=tree.getroot().xpath('//a:*',namespaces=NS)  # finds all the elements in the xml file
tags=[e.tag.split('}')[1] for e in elements]  # returns all the tags (not including the namespaces) of the elements
c=Counter(tags)  # a Counter object based on the tags list
print('A COUNT OF THE ELEMENTS IN THE XML FILE:')
print('; '.join(['{}: {:,}'.format(x, c[x]) for x in sorted(c.keys())]))  # formats and prints Counter c
print('')
df=pd.read_csv(csv)  # reads the csv file into a pandas DataFrame object
u=df['TimeSeriesVariable/@id'].unique()  # the unique values in the TimeSeriesVariable/@id column
print('A SUMMARY OF THE CSV FILE:')
print('Number of readings (i.e. rows): {:,}; Number of Variables: {:,}'.format(len(df),len(u)))
```

```
A COUNT OF THE ELEMENTS IN THE XML FILE:
Appliance: 618; Boiler: 20; BoilerControl: 20; Building: 20; Climate: 1; Cooker: 19; FixedHeater: 19; HotWaterCylinder: 13; Hous
ehold: 20; Light: 319; Meter: 40; Opening: 1,055; Person: 58; PhotovoltaicArray: 5; Plug: 421; Radiator: 252; RadiatorValve: 36
7; RefitXML: 1; RoomThermostat: 31; Sensor: 1,567; SolarThermalArray: 3; Space: 389; Stock: 1; Surface: 2,536; TimeSeriesVariabl
e: 2,457; WaterOutlet: 34

A SUMMARY OF THE CSV FILE:
Number of readings (i.e. rows): 25,312,397; Number of Variables: 2,320
```

Conclusions

- Open Science brings many benefits
  - Accelerate advancements in the field
  - Provide new researchers with existing methods to be improved upon
- Open Data should be well-structured and logical
  - json, xml, graphs, hdf5, object-orientated programming
  - it reduces the complexity of the analysis algorithms
- Tools exist to enable an Open Methodology workflow
  - GitHub, Python, Jupyter notebooks – but many others also exist
- A case study has been demonstrated
  - refitXML: not the only solution, but a start…
- Next steps
  - An Open Science journal paper
  - Open Science workflow for Building Simulation using EnergyPlus (…Python package needed…)

Thank you!

Dr Steven K Firth
S.K.Firth@lboro.ac.uk
https://github.com/stevenkfirth
REFIT Smart Home Dataset:
https://doi.org/10.17028/rd.lboro.2070091.v1

Loughborough University

Inspiring Winners Since 1909