

# Data Management for TEDDINET

DM4

**EPSRC** Data Management for TEDDINET  
using Semantic Technologies

Workshop 2  
Data formats and  
tools: how to  
enable data legacy

# Meeting practicalities

- Fire exit/alarm
- Bathrooms
- Breaks, lunch, dinner
- Travel expenses
  - <http://www.bath.ac.uk/finance-procurement/docs/visitors-expense-claim-form.pdf>
  - Or Google “bath university travel claim”
- Hotel(s)
- Round table(s) for introductions
- ... anything else?

# DM4T Aims + Context

- Enable a (unified?) data legacy for TEDDINET
- Help identify tools and techniques for STEM research to
  - curate data and
  - meet new RCUK requirements
- 18 months: March 2016 to September 2017
- Operate alongside TEDDINET/other EPSRC projects

# Project agenda

- Three phases:
  - Surveying current practice: essential vs. desirable
  - Dissemination of good practice: data management; ethical issues
  - Proof-of-concept tool prototypes: metadata creation; semantic tagging
- Lay foundations for
  - Integrated publication of data and articles
  - Semantic querying of diverse datasets

# Your data needs you!

Slide from subgroup session at the TEDDINET meeting in Nottingham at the end of April 2015

- institutional data archive - status? support?
- data access statements?
- data management policy?
- do you feel you know enough about what can do/should do?
- procedure to decide which data to preserve? and publish?
  - criteria for above?
- scale of data involved?
  - preservation plan for next 10 years?

Compliance deadline is **this Friday** 1st May 2015!

From summer 2015 EPSRC will start to monitor compliance!

# Data Management Challenges

- Legal requirements: familiarisation
- RCUK requirements: evolving
- Institutional implementation: evolving
- Little experience in STEM research... lots outside
- Data: raw? cleaned? format?
- Metadata: who? how? structural? descriptive?
- Access policy: origin? maintenance? constraints?
- Semantic tagging and querying?

# Future events

- Workshop 3 / 2016Q3:
  - ethics in practice
  - working with data (see next slide)
- Workshop 4 / 2017Q1:
  - joint workshop with TEDDINET, CEE
  - showcase tool demonstrators for data legacy
  - recommendations

# This meeting

**Aim for workshop 2:** to report on experiences with “next level” formats and tools applied to the energy research domain, with the overall goal of enabling the delivery of accessible data legacies for the TEDDINET portfolio.

Topics shortlisted were: RDF, ~~Mongo~~, HDF5, ~~JSON~~, ~~XML~~, ~~DTDs~~, Matlab, R and Python.



# Something to think about...

- An (anonymised) demonstration/training dataset?
- Different formats + different tools
- For example:
  - 1 week of elec+gas (if available) + sensors in one house
  - 1 month (or longer?) of elec+gas (if available) + outside temperature
  - 1 month (or longer?) of elec+gas use for two similar/adjacent homes (for a comparative analysis)

# Programme for today

13:15 - 14:15	Towards Frictionless Data	Jo Barratt, Open Knowledge Foundation
14:15 - 14:30	Overview of candidate data formats	Julian Padget, Bath
14:30 - 15:00	Data capture and storage in ENLITEN	Sukumar Natarajan, Bath
15:00 - 15:30	Hierarchical Data Format	Daniel Fosas De Pando, Bath
15:30 - 16:00	Coffee break	
16:00 - 16:30	Resource Description Format	Julian Padget, Bath
16:30 - 17:00	Data collection, storage and analysis in the APAtSCHE project	Bruce Stephen, Strathclyde
17:00 - 17:30	Lightning talks + wrap-up day 1	All