



# EUMETNET Crowdsourcing workshop

## Working groups

### **Theme A:**

# **Data sources, sharing and legal issues**

Facilitator: Lasse Latva

Duration: 20 minutes

15.02.18



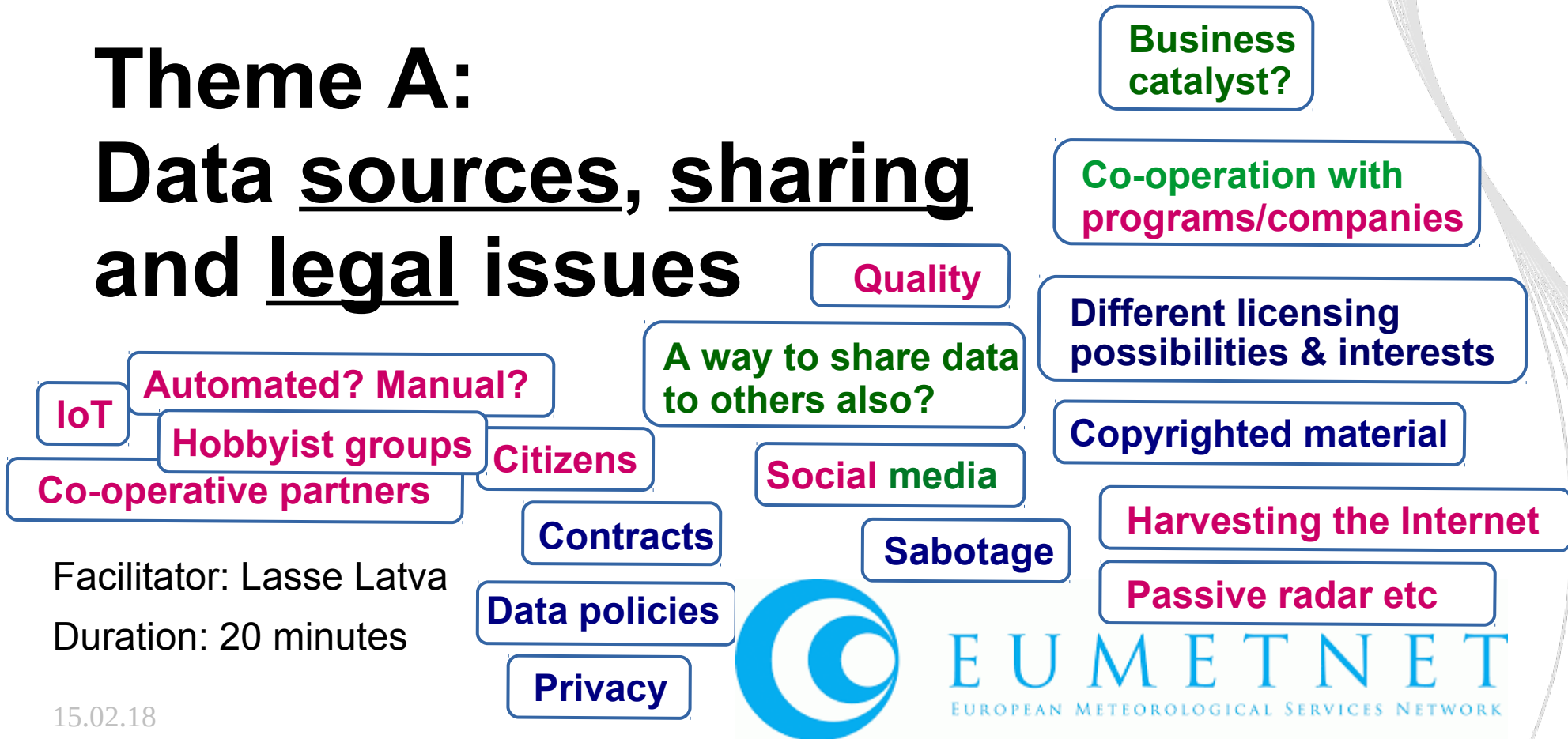
**EUMETNET**  
EUROPEAN METEOROLOGICAL SERVICES NETWORK



# EUMETNET Crowdsourcing workshop

## Working groups

### Theme A: Data sources, sharing and legal issues



Facilitator: Lasse Latva

Duration: 20 minutes





# EUMETNET Crowdsourcing workshop

## Theme A: Data sources, sharing and legal issues 1/2

### ***Data sources***

#### *Categorisation:*

- Meteorological data / impact data
- Amateur data / professional data
- Automatic data / manual data
  
- Updating and sharing (within EUMETNET) a list of data sources suggested!
  
- Lots of sources, most potential in PHONES, CARS (+DRONES), SMART CITY INFRASTRUCTURE (incl. also private solar cells, wind generators etc.)
  
- Big data mining; e.g. ice cream sales might correlate with weather
  - Automatically from photos rain/no rain, visibility, cloudiness – difficult for automatic stations
  
- Unique identifier is needed for each station/reporter



# EUMETNET Crowdsourcing workshop

## Theme A: Data sources, sharing and legal issues 2/2

### *Sharing & legal issues*

- Complex legal issues, contracts and privacy being probably most important areas
- Different actors have different requirements, e.g. car manufacturers not so willing to share data and e.g. insurance company's data must benefit that company itself. Companies will want licenses (and money) for their data.
- Privacy is a key issue
  - Requirements from legislation (EU GDPR)
  - Sometimes trade-off between data quality and privacy
  - Data privacy policies (and handling them in multi-service-provider environment)
  - Typically less sensitive data, user location being the most important
- Possibility of large-scale sabotage, e.g. false data ingestion (fake news..)





# EUMETNET Crowdsourcing workshop

## Working groups

### Theme B: Data formats

Facilitator: Roope Tervo

Duration: 20 minutes

15.02.18



**EUMETNET**  
EUROPEAN METEOROLOGICAL SERVICES NETWORK



# EUMETNET Crowdsourcing workshop

## Working groups

### Theme B: Data formats

Requirements for dataformats

Possible candidates

Quality information

Metadata handling

Standards

Software support

Facilitator: Roope Tervo

Duration: 20 minutes

15.02.18



**EUMETNET**  
EUROPEAN METEOROLOGICAL SERVICES NETWORK



# EUMETNET Crowdsourcing workshop

## Theme B: Data formats 1/2

- Used currently
  - GeoJSON, XML, Oregon, CSV, NetCDF, ASCII
- Requirements for data formats
  - Different format for different use cases (reports, sensors...)
  - Timely, readable and frequent is required sometimes
  - Should be easy to convert
  - Should be INSPIRE and WIGOS compliant
- Metadata
  - Elements (at least): quality, location & context (spatial and timely)
- Use cases
  - Data exchange (data hub), data collection, data usage





# EUMETNET Crowdsourcing workshop

## Theme B: Data formats 2/2

- How to get there?
  - EUMETNET working group
  - Guidelines for manufacturers (how to measure, model, encode and preprocess the data)
  - OGC one possibility
  - Collaborating with one or more big players (Google, Amazon, IBM...) could be useful
  - Can we use current formats and just flag them as crowd sourced
    - Develop current formats if needed







# EUMETNET Crowdsourcing workshop

## Working groups

### **Theme C:**

# **Image processing and quality control**

Facilitator: Ismo Karjalainen

Duration: 20 minutes

15.02.18



**EUMETNET**  
EUROPEAN METEOROLOGICAL SERVICES NETWORK



# EUMETNET Crowdsourcing workshop

## Theme C: Image processing and quality control 1/2

### Images in crowdsourcing

- Not that much of experience about automatic recognition, yet
  - Google Vision API and OpenCV was mentioned as well as machine learning in general
  - FMI has tried to use Flickr images, but way back when the smartphone cameras weren't too good
- Pictures and video clips are important for nowcasting!
  - Like webcams in general or road weather stations' cameras
- Pictures from the same places (weather stations) on good weather and bad weather (comparing between them and marking spots for e.g. visibility)
- Registering helps to prevent to get "bad/false" images
- Images help to gather metadata (surroundings of the stations)
- Manual checking is timeconsuming, but there are problems for automatic recognition of the images (size and shape of hail for example, but a reference object helps, e.g. a coin)
- We should use more public cameras from school, hospitals, government buildings etc.
- Smartphones can do the automatic recognition by themselves





# EUMETNET Crowdsourcing workshop

## Theme C: Image processing and quality control 2/2

### Quality control of crowdsourced data

- Mostly done manually now
- Using images help a lot
- Trained spotters are needed! Idea of different levels of them too
- Registering helps to get better observations, but most likely will affect on the number of observations
- We should inform groups, organisations, schools about how to do observations
  - They also may do some manual QC for us (we need to remember to thank them and give feedback!)
- Crowdsourced data can be used for verification (comparing them e.g. to synoptic obs)
- Can be compared with the weather models
- First observation is very important!
  - The forecaster can see, by his/her experience, that if the observation/image is good or not
  - The forecaster should get everything to see
- We should have campaigns, competitions ("best weather stations")

