WP1 work and priorities, DARE results and data storage issues in support of WP2

M Brunet on behalf of URV/C3 UERRA's team: L Ashcroft, JR Coll, M Castellà, P Domonkos, E Aguilar, J. Sigró and 11 geography students (I López, E. Romero, J Paradinas, A. Robert, J Tarragó, A Domènech, E Pla, C Bonfill, G Vandellòs, R Guerra, D Azuara + Y2 11 students: I López, J Paradinas, A Àvila, A Balart, M Martin, E Cebrián, D Azuara,P Sabaté, A Boqué, A Domènech, E Pla)

UERRA WP1 objectives and tasks: an overview

Work package number [53]	WP1	Type of activity [54]		RTD
Work package title	Data Rescue and development, gridded and obervational datasets			
Start month	1			
End month	36			
Lead beneficiary number [55]	7			

- T1.1 - Data coordination, inventory and access to national archives [Months: 1-36] – URV (3.7M hourly observations), NMA-RO (~ 300K) D1.1 to D1.4

- T1.2 - High-quality synoptic-scale data development [Months: 1-36] - URV, UEA D1.5 to D1.9 and D1.13

- T1.3 - Gridded and Observational Datasets [Months: 1-36] - KNMI, UEA, EDI D1.10 to D1.12 and D1.14
  - Enhance gridding procedures within E-OBS, particularly for extremes
  - Improve the uncertainty assessment within E-OBS

WP1 deliverables list, the focus for Y1 and its status

				Estimated...				
D1.1	DARE list of sources	WP1	7	4.00	Report	PU		6
D1.2	DARE station locations	WP1	7	2.00	Report	PU		10

T1.1 - Data coordination, inventory and access to national archives
- Identifying the DARE targets: exploring digitised data availability from MARS (CRU/UEA and C3/URV) and from the gathered (EURO4M) and new data-sources in scanned format
- Exchange exercise (scanned data by digitised and QC'ed data) proposed to several NMS
- Digitisation committed (3.7M and 3K stations values by URV & NMA-RO), its status and progress: URV ~3.5M values – NMA-RO: 3 RR stations for 1979 - 2002
- Two first deliverables provided on time

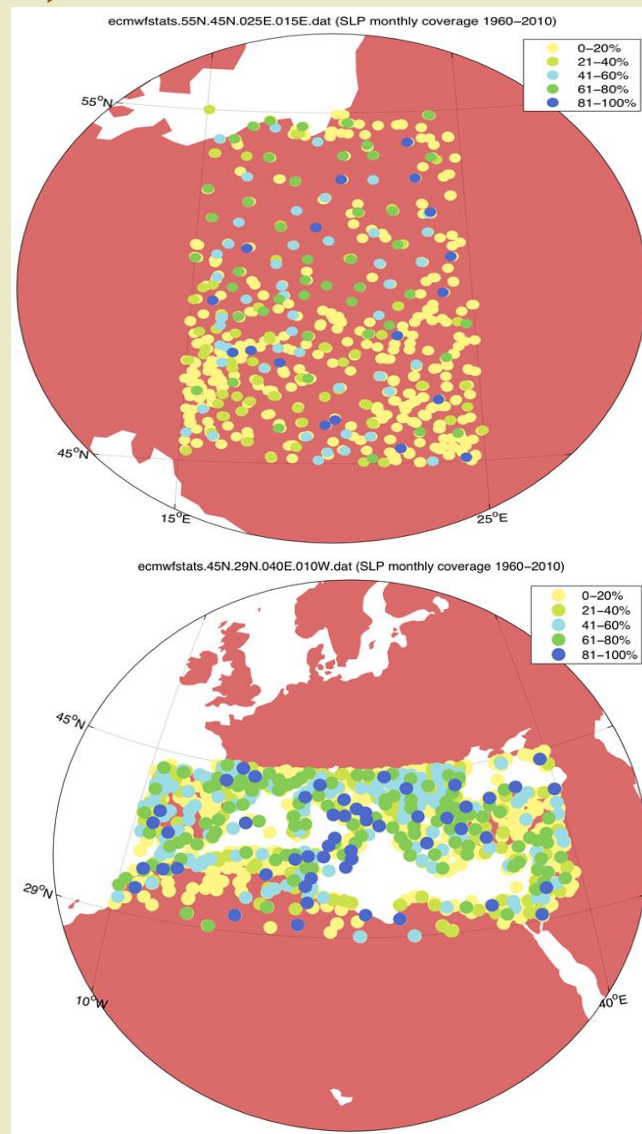**T1.2 - High-quality synoptic-scale data development**
- Defining and implementing new QC to test: a double strategy VQC and AQC (ensuring data sources and digitised data quality). Starting QC application and need for auxiliary data (Linden on this later)
- CRUTS dataset updating to 2013, extending to Europe CRUTEM global temperature archive and merging some ECA&D stations and CRUTEM at the monthly scale (by CRU/UEA)

T1.3 - Gridded and Observational Datasets
- Improving the Gridding of precipitation in the E-OBS Dataset (Phil on this later)
- Continuing to release monthly updates of daily gridded station data for Europe, E-OBS, and preliminary work towards deliverable D1.9: Assessment of the impact of changes in station density on the E-OBS dataset by KNMI has started (Albert on this later)
- Quantifying interpolation uncertainty via stochastic simulation, aimed at deriving an ensemble of quasi-realistic grids of precipitation conditioned on rain-gauge observations and representative for the uncertainty inherent to the limited station density by EDI

Setting the DARE targets, digitisation progress and caveats (T1.1)

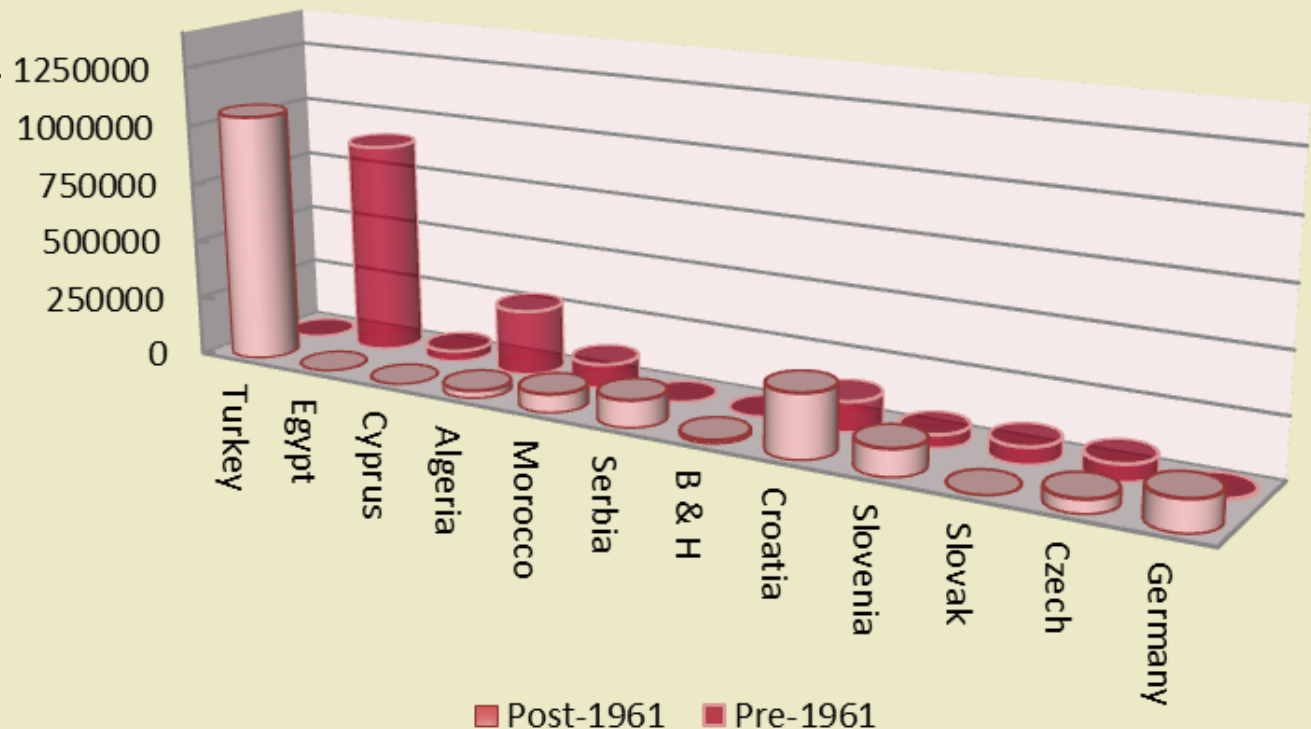- Defining digitisation targets: survey to UERRA relevant partners & meeting at ECMWF to set targets:
  - Variables: SLP, TMP, WS, WD, RH, DP, SD, FS, RR at the hourly and daily scales
  - Post-1961 (the target), but also pre-1961
  - Southern part of Med, Eastern Europe & Balkans, but subjected to data sources availability
- Exploring MARS (CRU) & crosschecking (URV) with gathered imaged data-sources (from EURO4M), data coordination (NCDC new holdings, ISPD, ISTI-DARE) and accessing national archives
  - New data sources added to those gathered under EURO4M. Data-sources for pre-1961 obs easy to find/access, but difficult for post-1961 data. Then…
  - Exchange exercise (undigitised data by digitised data) proposed to Catalonia, Egypt, Jordan, Libya, Germany, Macedonia the FYR, Montenegro, Romania, Serbia, Slovenia and Sweden NMS. Positive responses from most of them (e.g. Jordan, Libya, Macedonia the FYR, Montenegro, Serbia) and the negative of others (e.g Romania and Sweden), but only proposals to Catalonia, Germany and Slovenia have materialised so far. Reasons: from internal data digitisation policies (Romania, Sweden) to difficulties of other NMS to have scanned - duplicated their recent obs.

ecmwfstats.55N.45N.025E.015E.dat (SLP monthly coverage 1960–2010)

0–20%
21–40%
41–60%
61–80%
81–100%

ecmwfstats.45N.29N.040E.010W.dat (SLP monthly coverage 1960–2010)

0–20%
21–40%
41–60%
61–80%
81–100%

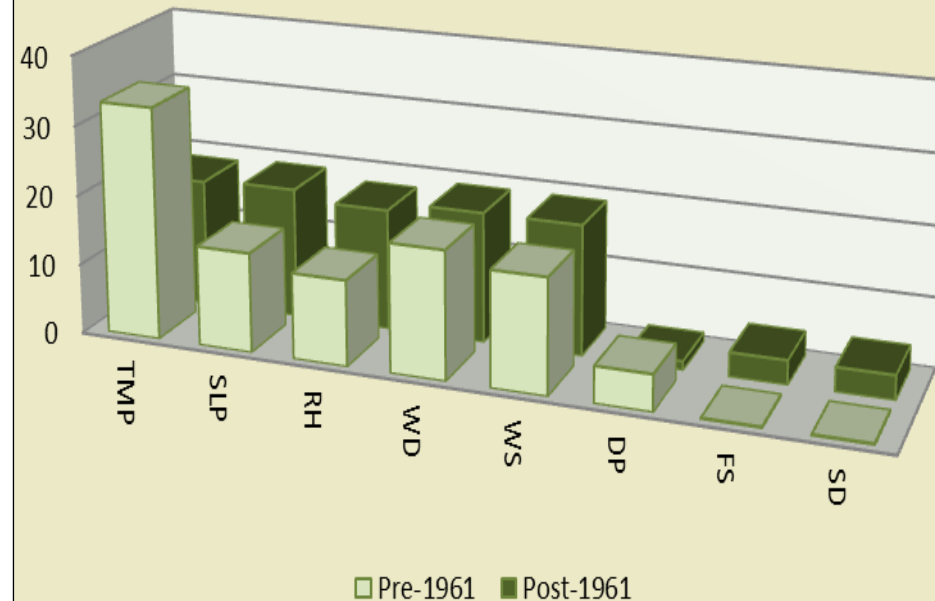# URV digitisation progress over Y1 & status (T1.1) I

- EURO4M SLP data reformatted to ODB (CRU) and provided to MARS, but in absence of guidelines the MARS convention for date and time columns couldn't be followed and CRU will have to reformat/provide the dataset again

- Focus on post-1961 hourly obs., but also recovered pre-1961 values due to initially limited data sources availability

- About 3.5M station-values already digitised: 1.8M (1.6M) for post-1961 (pre-1961)

- Remarkable for Turkey and Balkans (Egypt, Algeria) areas for post-1961 (pre-1961)

- Effort over eastern Germany SD and FS post-1961 obs.

No. digitised data by countries and for post- and pre-1961 periods: 3.5M station-values

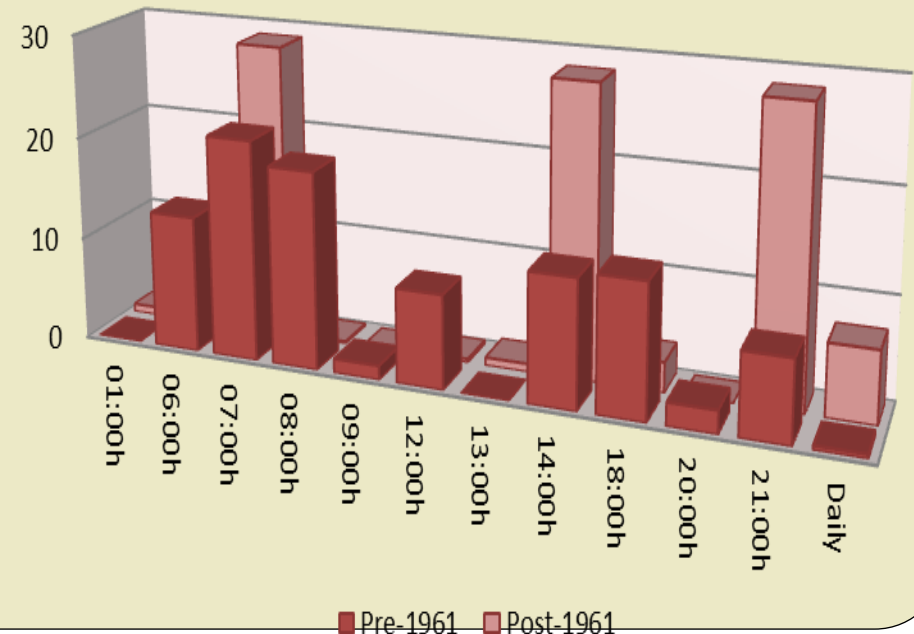Post-1961   Pre-1961

URV digitisation progress over Y1 & status (T1.1) II

% digitised values by variables and pre- (1.6M) and post-1961 (1.8M)

Pre-1961   Post-1961

More diversity times digitised for pre-1961, while for post-1961 3 obs. times/day predominate (7h, 14h & 21h)

% digitised values by observing time and pre-1961 (1.6M) amd post-1961 (1.8M)

Pre-1961   Post-1961

More balanced digitisation among variables for post-1961, but DP and FS/SD, when compared with pre-1961: TMP, followed by WD/WS and SLP and RH in less extent (little or nothing for FS/SD) digitised

Data storage issues in support of WP2

- Remarkable amount of data already digitised, but still pending of being completely QC'ed to be delivered. This is the focus for Y2 (Linden on our QC status & approaches)

- Data storage issues:
  - EURO4M hourly SLP and daily Tx, Tn and RR data still sitting outside of ECMWF (at ZENODO: https://zenodo.org/record/7531, ECA&D, ISTI and ISPD databanks + CRUTEM for monthly Tx/Tn)
  - EURO4M SLP data because have to be converted to ODB again. Are daily Tx, Tn and RR data in need for UERRA to be provided to MARS after conversion in ODB by CRU?

- ECMWF updates MARS once per year. This could be an issue for using the recovered data for WP2, although UK-MO production runs will not start before 2016. However, SMHI need to start earlier

- If current digitised data (and EURO4M daily observations) are also of interest for any component of the UERRA work-load, it should be defined both the data format and the delivery mechanism that suits better with the relevant UERRA WPs.

- Any thoughts on data format and delivery mechanisms, in the case MARS annual updates are not frequent enough ?

Discussion & suggestions on data format & delivery mechanism welcome