Reply To The Editor

Author Response:
We would like to thank the reviewer for taking the necessary time to review the manuscript. We are grateful for all the valuable comments, which helped us to improve the quality of the manuscript.

Responses to the comments in the PDF supplement to Anonymous Reviewer #1 (Report #1)

Line 92:
Highlighting of “.”

Reviewer Comment:
extra full stop

Author Response:
We agree with the reviewer.

Author Change:
Removed the extra full stop.

Line 103:
Highlighting of “during”

Reviewer Comment:
should be 'during-,'

Author Response:
We agree with the reviewer.

Author Change:
“As data documentation during observation campaigns occurs at pre-, during- and post-deployment, it is critical to have a structure early for capturing all details (e.g., Table 1), especially at busy, time-limited periods.”

Line 107:
Highlighting of “it”

Reviewer Comment:
use 'data' to be more precise

Author Response:
We agree with the reviewer.

Author Change:
“Critical to this, are data being accessible to team members during a campaign and subsequently as data are processed (Figure 2).”

Line 118-121:
Highllighting of “e.g.,” and “plan”

Reviewer Comment:
this seems unnecessary

Reviewer Comment:
plan, and

Author Response:
We agree with the reviewer.

Author Change:
“Those help find (1) if an identical instrument model exists in storage or in another deployment, (2) all instruments at a location, (3) all mobile-phone SIM cards linked to a data plan, and (4) all calibration sheets and warranty documents for a given sensor to facilitate sending an instrument back to a manufacturer for service.”

Line 123:
Highlighting a section of the line.

Reviewer Comment:
'information about an instrument's'

Author Response:
We agree with the reviewer.

Author Change:
“The Deployment DB (Figure 3) has information about an instrument’s configuration, including location and relation to other instruments during a deployment, as well as organisation details.”

Line 147-150:
Highlighting of VDI

Reviewer Comment:
What is VDI and why is it just an example. What is the overarching convention for automated assessment of data?

Author Response:
We agree with the reviewer that this line is confusing.

VDI is an undefined acronym in this line, but also a citation and an example of a standards authority. There is currently no overarching convention that we are aware of that would cover the QC procedures applicable to data from all system groups listed in Figure 1 and Table 3.

The section about quality control was moved to an appendix during an earlier revision, making the undefined acronym and reference for VDI in the main text inappropriate. We propose to keep the AC section in Appendix B and change the line.

Author Change:
"Data stream quality control (QC) includes automated assessment of typical meteorological variables (e.g., air temperature, humidity, wind speed, wind direction, pressure, precipitation intensity), following convention and standardized procedures (see Appendix B)."

Line 155:
Highlighting of "API"

Reviewer Comment:
Has this been defined?

Author Response:
We agree with the reviewer's observation that the acronym "API" is not defined before this line in the text.

Author Change:
The definition for API was added here and removed later in the text (Section 32. Access).

Line 193-194: "All systems have GPS time or internet reference time services, and all data during urbisphere campaigns are recorded in UTC or the GMT/UTC+0000 locale without daylight saving for systems with time-zone unaware recording of timestamps (Appendix A1)."

Reviewer Comment:
define "UTC"

Reviewer Comment:
define "GMT"

Author Response:
We agree with the reviewer.

Author Change:
"All systems have GPS time or internet reference time services, and all data during urbisphere campaigns are recorded in Coordinated Universal Time (UTC), Greenwich Mean Time (GMT) or the GMT/UTC+0000 locale without daylight saving for systems with time-zone unaware recording of timestamps (Appendix A1).

Line 214-216:
“The core database systems, including the inventory and deployment configuration DBs (Figure 3), are designed using open source web-, database- and user-interface tools (the so-called LEMP stack; Linux, Nginx, MariaDB, PHP) and application frameworks (Appendix C3).”

Reviewer Comment:
previously these were referred to as Inventory DB and Deployment DB, this terminology is inconsistent with prior uses.

Author Response:
We agree with the reviewer.

Author Change:
“The core database systems, including the Inventory DB and Deployment DB (Figure 3), are designed using open source web-, database- and user-interface tools (the so-called LEMP stack; Linux, Nginx, MariaDB, PHP) and application frameworks (Appendix C3).”

Line 237: “The original data and metadata are kept on different physical servers. However, users with access to a workstation are provided with immediate access to a read-only view of the original data, as well as a read-only replica of the metadata DBs.”

Reviewer Comment:
which DBs are these? all of them?

Author Response:
We agree with the reviewer that this is unclearly formulated.

Author Change:
“The original data and metadata are kept on different physical servers. However, users with access to a workstation are provided with immediate access to a read-only view of the original data, as well as a read-only replica of the metadata DBs (Section 2.1).”

Line 241:
Reviewer Comment:
This is the third occurrence of API and should be defined at the first occurrence.

Author Response:
We agree with the reviewer.

Author Change:
The definition has been moved to the first occurrence in the main text.
“Workstations are used as public access nodes, with a wide selection of services available at the user-level, including APIs, Integrated Computing Interfaces (ICEs) and other interactive websites (Apps) for users and the public (see also Appendix C).”

Line 247-248:
“The need for administrator privileges is avoided, where possible, and the directory access for accounts used in automation (e.g., FTP credentials are transmitted as text) is restricted in scope.”

Reviewer Comment:
define “FTP”

Author Response:
We agree with the reviewer.

Author Change:
“The need for administrator privileges is avoided, where possible, and the access for accounts used in automation (e.g., File Transfer Protocol (FTP) credentials are transmitted as text) is restricted in scope.”

Line 281:
Highlighting of “NBIoT”

Reviewer Comment:
What is this? Internet of Things was mentioned previously, but the acronym was not given. And NBIoT is a new term.

Author Response:
We agree with the reviewer.

Author Change:
The acronym for Internet of Things was introduced where it was mentioned previously, and the line was changed to include “Narrowband IoT (NB-IoT) network services” instead of “NBIoT”.

Line 288: “SFTP adds a secure authentication and encryption layer to the transfer (cf. FTP), whereas RSYNC adds incremental, compressed and validated data transfer (cf. SFTP).”

Reviewer Comment:
what does cf. stand for?

Author Response:
This is a common abbreviation for the Latin ‘confere’, meaning ‘compare’.

Author Change:
No change to the text.

Line 288-294:
“The logical network uses industry-standard protocols for the transmission of data files (i.e., FTP, SFTP and RSYNC through SSH). SFTP adds a secure authentication and encryption layer to the transfer (cf. FTP), whereas RSYNC adds incremental, compressed and validated data transfer (cf. SFTP). RSYNC is preferred, as it allows reliable recovery of incomplete or failed transfers with limited bandwidth overhead on the logical network. Custom software is used to configure the RSYNC client software and set retention periods for data transmission (Morrison, 2022). The synchronisation of data between storage locations also relies on RSYNC (i.e., as transport method for Lsyncd). We find the FTP protocol is no longer fully supported by all mobile phone network carriers. As some data loggers (e.g., model CR1000X, Campbell Scientific, Logan, Utah, USA) use alternative protocols, the upload server is configured to allow legacy authentication methods for SFTP connection.”

Reviewer Comment:
please describe RSYNC

Reviewer Comment:
what is this? [lsyncd]

Author Response:
We agree with the reviewer. These lines were imprecise in the reference and description of the software utilities Rsync and Lsyncd, requiring technical editing.

Author Change:
“The logical network uses industry-standard protocols for the transmission of data files (i.e., FTP, Secure FTP (SFTP) and Secure Shell (SSH) in combination with the Rsync network file transfer software). SFTP adds a secure authentication and encryption layer to the transfer (cf. FTP), whereas the Rsync software adds incremental, compressed and validated data transfer (cf. SFTP). Rsync is preferred, as it allows reliable recovery of incomplete or failed transfers with limited bandwidth overhead on the logical network. Custom software is used to configure the Rsync client software and set retention periods for data transmission (Morrison, 2022). The synchronisation of data between storage locations also relies on Rsync (i.e., as transport method for the Lsyncd file synchronisation software). We find the FTP protocol is no longer fully supported by all mobile phone network carriers. As some data loggers (e.g., model CR1000X, Campbell Scientific, Logan, Utah, USA) use alternative protocols, the upload server is configured to allow legacy authentication methods for SFTP connection.

Line 304: “Participants and data systems produce many data set and services.”
Reviewer Comment:
I believe this [set] should be plural

Author Response:
We agree with the reviewer.

Author Change:
“Participants and data systems produce many data sets and services.”

Line 313: highlighted word ‘Metadata’
Reviewer Comment:
Check that both metadata and meta data are not used in the manuscript.

Author Response:
We agree with the reviewer and checked again that only ‘metadata’ is used, for consistency.

Author Change:
No changes made.

Line 320: highlighted interactive computing environments (ICE) in “Web-based interactive computing environments (ICE) on workstations with common libraries and replicate environments are to develop code with immediate access to the data archive.“

Reviewer Comment:
this is not the first occurrence, please define at first occurrence.

Author Response:
We agree with the reviewer on this technical omission.

Author Change:
“Web-based ICE on workstations with common libraries and replicate programming environments are to develop code with immediate access to the data archive.”

Line 329-330: highlighting “NAGIOS” in “A software tool monitors system status, resource monitoring and alerts (“watchdog”) using the NAGIOS protocol and tools.”

Reviewer Comment:
what is this?

Author Response:
We agree with the reviewer that this line is unclear.

Author Change:
“The monitoring of computer system status, resource use and alerts (``watchdog") uses the open-source Nagios protocol and software.”

Line 331-332: highlighting “Application programming interfaces” in “Application programming interfaces can enhance data access by providing dedicating handling of communication between computer programs and are used for many tasks.”
Reviewer Comment:
APIs

Author Response:
We agree with the reviewer.

Author Change:
“APIs can enhance data access by providing dedicated handling of communication between computer programs and are used for many tasks.“

Line 359-360: Highlighting “partners” and “they contribute to.” in “Most campaigns use data streams from partners instrumentation either directly or more typically from their data networks (e.g.,Weather Service), or from third-party networks they contribute to …”

Reviewer Comment:
not plural

Reviewer Comment:
do not end a sentence with a preposition

Author Response:
We agree with the reviewer.

Author Change:
“Most campaigns use data streams from partner instrumentation either directly or more typically from their data networks (e.g., Weather Service), or from third-party networks …”

Line 363: “Examples of different type of roles of users of the system include …”

Reviewer Comment:
of user roles for

Author Response:
We agree with the reviewer.

Author Change:
“Examples of different type of user roles of the system include …”

Line 391: Highlighting “URI”
Reviewer Comment:
what is this?

Author Response:
We agree with the reviewer that URI was not defined.

Author Change:
Replaced “URI” with “Uniform Resource Identifier (URI)”

Table 1:
Highlighting of the caption

Reviewer Comment:
What is the main point of this table. There should be a short description of the overarching point. I would also recommend spending some more time in the text explaining this table and how to interpret it with this example

Author Response:
We agree with the reviewer that the text and caption need to be changed to clarify the main point of the figure.

Author Change:

We propose the following changes.

Caption "Table 1. The provenance of each field observation can be mapped within an infrastructural-, a logical- and an organizational network. The connections and associations between the origin and the data product (both in bold) are not limited to the field situation (indicated by asterisk)."

The presentation of this Table in the text, with help of Figure 2 and Figure 5 in Section 3, is updated to:

"A "data source" may be a sensor, a network node or an organisational unit (Table 1), with different contexts that need to be retained and clearly identifiable. Typically, a chain of systems and responsibilities are involved, with multiple actors, nodes and locations (Table 1; Table 3; Figure 5). The origin of data may be expressed in terms of the physical network of infrastructure at distributed locations, the logical network involved in data telemetry, storage and processing, and the network of organisations and actors who have various roles. The source needs to be defined and preserved in order to ensure data governance agreements, accessibility, responsibilities, and also to effectively respond to issues that occur. For example, if an instrument at a particular location is not responding, the data and metadata must allow the relevant infrastructure, the responsible people and the production line processes to be looked up efficiently for that particular source (Figure 2)."

Table 4:
Highlighting of the caption.

Reviewer Comment:
I recommend adding horizontal lines as this is really difficult to read

Author Response:
We agree with the reviewer.

Author Change:
Horizontal lines were added to separate the Modules.

Table 5:
Highlighting of the caption.

Reviewer Comment:
Also add horizontal lines here

Author Response:
We agree with the reviewer.

Author Change:
Horizontal lines were added to separate groups of Terms.

Table 6:
highlighting the “Creator” row.

Reviewer Comment:
this seems to not fit into just one of the above columns. Is this production or publication?

Author Response:
We agree that this is unclear. The “Creator”, “Material”,”Attribution” and “Disclaimer” texts are the (nearly) the same for Production and Publication.

Author Change:
A horizontal line was added to visually separate the top and bottom items.

Figure 4:
highlighting “report data if powered, so metadata are needed to define operational deployment periods.”

Reviewer Comment:
Is this because they might be powered but not gathering data on the campaign? Why not just turn off the power or data transmission when not in use?

Author Response:
This is exactly right, but not easy to implement. Both the power and data transmission (LTE or NB-IoT) are designed to be inaccessible and always-on, as these systems are deployed in public space on street lights. A battery is fixed in a weather-sealed enclosure, which is closed in preparation of deployment/calibration. The batteries can last for months. During maintenance, systems are replaced (hot-swapped) where needed and a maintenance protocol (in paper and digital) informs on the modification made by the technical staff.

Author Change:
No change made.

Figure 5:
Highlighting of “including two automated methods (M1, M2).”

Reviewer Comment:
It would be good to spend some more time describing the difference between the two automated methods and why the user may choose one over the other.

Author Response:
We agree with the reviewer that this difference is clearly specified here. We are working on an evaluation of automated procedures, including the skill improvements for QC by using novel signal analysis techniques compared to the current standard. The difference between the conceptual M1 and M2 methods can be best outlined in a separate publication.

Author Change:
No change to the text.

Table A1:
Highlighting the caption

Reviewer Comment:
add in horizontal lines

Author Response:
We agree with the reviewer.

Author Change:
We have added horizontal lines in order to visually separate “station”, “system”, “sensor” and “cell” names.

Line 515, Line 535, Appendix A1:

Reviewer Comment:
It would be good to include a mention about denoting the beginning and end of a time interval

Reviewer Comment:
I would recommend spending some more time about this. It is really important to give the start and the end of an averaging or data acquisition period, and that is not the default for many systems.

Author Response:
We agree with the reviewer that it is important to give the boundaries of (any) an interval. We propose to include a reformulation of our earlier response to an anonymous review comment

pointing out this importance. The cf-conventions (CF) provide guidelines on time, and so does the ISO8601 standard.

Author Change:
We have added text about the importance of time notation in data and metadata to Appendix A1.

“An accurate time convention is critical. The CF convention offers comprehensive and clearly defined options to describe and encode the start and end of intervals of time (or any other dimension). A basic option is to declare the relevant attributes of the "cell" value. For many of the data products, the data are "point" samples stored in their original sampling resolution. This is specified as an attribute to the variable (e.g., cell_method: "time: point"). Upon any aggregation, along the time dimension or any spatial dimension, the bounds can be (must be) declared and the cell_method attribute for those variables updated accordingly. The CF convention further reserves suffixes for aggregation to be added to the variable name, such as _mean and _maximum, to indicate aggregation of data has occurred. Adding the time bounds technically implies adding a 2-position virtual dimension to the data structure, in order to store both the start and end of the interval coordinates along dimension time.

Besides the use of time bounds for aggregation periods in the data, also metadata may require intervals to be specified. The ISO8601 standard provides a detailed formatting description for the representation of periods as a machine readable text. This notation works well in metadata records, including attributes in NetCDF data files, but the separator between the start- and
end-time is defined as a forward slash character and is incompatible with use of in file names (see Section B1).”

Line 541:
Highlighting ‘Verein Deutscher Ingenieure (VDI)’.

Reviewer Comment:
This needs to be added to the main text along with a short summary of what it is.

Author Response:
We agree with the reviewer that a solution is needed here and in Line 147-150.

Author Change:
See proposed changes to Line 147-150.

Responses to the comments by the editorial office / Remarks from the preceding review file validation

Review Comment:
“Your reference list includes works “in preparation” or “in review”. Such works can be cited upon submission if being available to the reviewers. They cannot be cited in the final, accepted manuscript, unless published, accepted for publication, or available as preprint with a DOI.”

Author Response:
The work this applies to,
Feigel, G., Plein, M., Zeeman, M., Metzger, S., Matzarakis, A., Schindler, D., and Christen, A.: High spatio-temporal and continuous monitoring of outdoor thermal comfort in urban areas: a generic and modular sensor network and outreach platform, Sustainable Cities and Society, in review.

Author Change:
A preprint solution was requested for this work, but it is not yet available. If this preprint is not available on time for the final accepted manuscript we propose to replace the reference with:

Feigel, G., Plein, M., Zeeman, M., & Christen, A. (2023, October 13). Realtime Weather Monitoring Network Freiburg. AK Klima, Geographisches Institut der Universität Tübingen. Zenodo. https://doi.org/10.5281/zenodo.10549973