



SOFTWARE

Open Access

# A software tool for the input and management of phenotypic data using personal digital assistants and other mobile devices

Karin Köhl\* and Jürgen Gremmels

## Abstract

**Background:** Plant breeding and genetics demand fast, exact and reproducible phenotyping. Efficient statistical evaluation of phenotyping data requires standardised data storage ensuring long-term data availability while maintaining intellectual property rights. This is state of the art at phenomics centres, which, however, are unavailable for most scientists. For them we developed a simple and cost-efficient system, the Phenotyper, which employs mobile devices or personal digital assistants (PDA) for on-site data entry and open-source software for data management.

**Results:** A graphical user interface (GUI) on a PDA replaces paper-based form sheet and data entry on a desktop. The user can define his phenotyping schemes in a web tool without in-depth knowledge of the system and thus adjust it more easily to new research aspects than in a classical laboratory information management system (LIMS). In the Phenotyper, schemes are built from controlled vocabulary gained from published ontologies. Vocabulary and schemes are stored in a database that also manages the user access. From the web page, schemes are downloaded as extended markup language (XML) files for the transfer to the PDA and the exchange between users. On the PDA, the GUI displays the schemes and stores data in comma separated value format and XML format. After manual quality control, data are uploaded via a web page to an independently hosted results database, in which data are stored in an entity-attribute-value structure to provide maximum flexibility. Datasets are linked to the original and curated data files stored on a file server. The ownership stamp, project affiliation and date stamp of a dataset are used to regulate data access, which is restricted to data belonging to the user or to his projects and data, for which the embargo period has ended. By export of standardised ASCII reports to long-term data storage facility, long-term accessibility allows searching, citing and use of raw data beyond the lifetime of the database. The Phenotyper is available to the scientific community for use and further development.

**Conclusions:** The Phenotyper provides a well-structured, but flexible data acquisition and management structure for mobile on-site measurements for efficient evaluation and shared use of data.

**Keywords:** Phenotyping, Field research, Relational database, Data management, Controlled vocabulary

## Background

Evaluation of plant genotypes in breeding programs or genomics projects require precise and reproducible phenotyping [1]. The coining of the term 'phenomics' indicates that morphological and physiological data are as important as biochemical and molecular information [2]. The lack of phenotype information rather than the lack of genetic information causes a bottleneck in generating

the genotype-phenotype map that links genes to relevant features like fitness or survival [3,4]. With respect to the amount of publicly available data, phenomics lags behind genomics [4]. Large genomics projects have been pioneers in making data rapidly available to scientific communities on web pages. Large research centres ([5-10]) phenotype plants highly reproducible under standardised conditions and store information in databases. These systems are, however, expensive, as infrastructure for data collection and storage require high financial and personnel investment [11]. These phenomics facilities are thus unavailable for most of the scientific community, which

\* Correspondence: koehl@mpimp-golm.mpg.de  
Max-Planck-Institute of Molecular Plant Physiology, 14424 Potsdam-Golm, Germany

nevertheless produces an increasing amount of valuable phenotypic data. The limiting factors for the optimal use of these data within models that link genetic, environmental and phenotypic data are data management and analysis [12,13]. Most scientists combine phenotyping with conventional methods of data recording and management. Especially in field-based research, data are frequently recorded in individual formats on paper, often without a safety copy. During manual entry into the computer for further evaluation errors are introduced into the dataset. In plant sciences, data are generally entered, stored and evaluated in spreadsheet programs like MS Excel (Microsoft, Redmond, US). Data storage in Excel is considered risky as data are easily compromised by unintentional use of automatic formatting function [14]. Furthermore, raw data are often mixed up with evaluation procedures and results, including figures. Column headings that indicate the content of a column are generally unstandardised leading to different names for the same parameter or the same name for parameter that differ e.g. in the unit. Furthermore, data are often spread to several files, the connection between datasets is ambiguous and the file storage structure determined by the personal preferences of the individual researcher. Access is often limited to the individual researcher. Frequently, data are deleted after the researcher left the institution and even if the data are archived, insufficient documentation of individual file names and column heads make the information ambiguous and impede searches. All these facts render the compilation of a complete set of raw data as basis of a comprehensive evaluation or to provide original data when publishing a manuscript very time-consuming. In phenotyping, this is especially deplorable. Many phenotypic features are semi-quantitative (colour, shape, developmental stage) [15,16] or highly variable (growth rate) and thus require large datasets to gain sufficient statistical power. Therefore, the combination of phenotypic data from many experiments and several users is a prerequisite for a successful project. Data from various sources should therefore be available rapidly for automatic evaluation. This can be achieved by using a central data repository with a highly standardised data storage structure based on publicly available controlled vocabulary. Several initiatives offer central data storage and sharing structures (overview see [11]). The first phenotype databases concentrated on data generated by standardised methodology that results in well-defined data structures like data from transcriptomics or ionomics [17-19]. In the American cyberinfrastructure project [20], the PODD Data Repository Project in Australia [21,22] and the Phenoscape project [23], information of highly variable structure are stored as free text descriptions enhanced by a semantic search algorithms that link texts with ontological terms, thus providing the

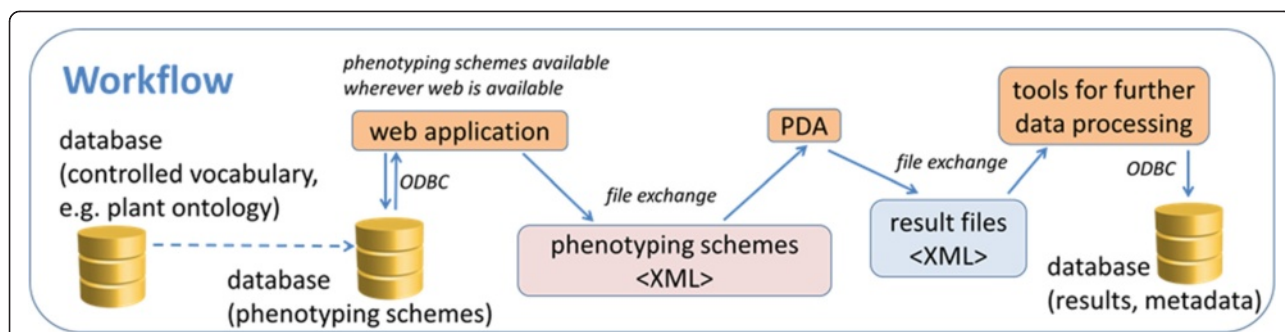
basis for efficient search functions. In addition to raw data, data analysis procedures need to be documented and made available to convert data efficiently into results [4] and to provide a reproducible connection between raw data and published results [24]. Within plant phenotyping, image analysis is furthest advanced with respect to automated data storage and analysis [8,25,26].

To facilitate the generation of large, centrally stored phenotyping datasets from manually or semi-automatically collected data, we generated a simple and inexpensive tool, the Phenotyper that uses mobile devices for on-site data entry. The system allows tailoring a graphical user interface (GUI) to the individual project in a web-based building-tool without programming skills or database knowledge. Phenotyping results are transferred as XML files via an internet page into a central database, from where they can be downloaded in user-defined, access-controlled mode by web tools.

## Results and discussion

### Phenotyping schemes based on controlled vocabulary from ontologies

The core feature of the phenotyping workflow is the combination of controlled vocabulary for entities, e.g. plant organs, and attributes, e.g. features that are to be scored, in user-defined phenotyping schemes. These are exported to mobile devices and displayed in a GUI (see Figure 1). By using controlled vocabulary, it achieves a high degree of standardisation during data recording while maintaining flexibility. Controlled vocabulary is used directly or indirectly by several data management systems for phenotypic data (e.g. the Triticeae Toolbox [27], Phenoscape [23], DbNP [28]) as it makes search and automatic analysis by computational algorithms more efficient [11]. The phenotyping schemes are composed on a web page (see Figure 2A), that retrieves the controlled vocabulary from the scheme database (S-DB) (Entity relationship (ER) diagram see Additional file 1: Figure S1). The controlled vocabulary was derived from the Plant Ontology (PO) Consortium [29] for entities. For parameters, it was partially user-defined, partially gained from the plant trait ontology (TO). By storing PO or TO identifiers, the connection towards the original source was maintained. The definition of project-specific vocabulary subsets improves the handling of the Phenotyper. The recording of schemes in a user-specific mode avoids accidental modification of the scheme by other users. Schemes are exported in XML format to be displayed in the GUI of mobile devices (see below). Cloning of schemes on the web page supports efficient scheme modification. Exchange of schemes between users is performed by importing scheme XML files into the S-DB. A language switch function, which changes the language on the web page, for the controlled vocabulary



**Figure 1 Workflow of the phenotyping system.** Phenotyping schemes are built in web applications from controlled vocabulary, which is stored in the S-DB (left), and submitted to the mobile device (personal digital assistant PDA). After phenotyping, results are transferred by web tools to the R-DB (right).

**A Manage phenotyping schemes**

Karin Koehl     English  Deutsch    [Go to administration page.](#)

logout    Change language

Test program name: LEO\_FW    Version: 17112014

Object: Aliquot    Test object type: Plant    Description:

Entity: shoot    Parameter and value: absolute freshweight g    Delete    Order number: 0

Add parameter value combination

New test program    Clone current test program    Cancel    Project: (no project selected)

Export test program    Save test program    Update test program

---

**B TPC Administration Page**

Entities    Attributes and Values    Projects


Attribute and value: anthocyanin high    Update    New

Attribute (EN): anthocyanin

Attribute (DE): Anthocyanin

Value (EN): high

Value (DE): high



copyright: Karin Koehl

Upload image

Back to main page

**Figure 2 Web tool phenotyping scheme composer and administration page. A.** The web tool combines controlled vocabulary of entity and attribute-value information in user-defined schemes that are stored in the S-DB and exported to the mobile device. **B.** Controlled vocabulary is managed on the Administration pages, which also provide an upload function for illustrating image files.

and on the mobile device, facilitates the use by technical personnel in multinational projects.

The user management of the S-DB also handles the access management to the web page. Users with administrator status can access additional pages, on which they can create new user accounts and projects, affiliate users to projects, define project-specific vocabulary subsets, introduce additional or modify existing vocabulary and upload picture files to illustrate categorical parameters (see Figure 2B). These tasks can be thus performed by project leaders without in-depth database knowledge independent of the database administrator. The ease with which the Phenotyper can be adjusted to the needs of an individual project and the open-source software model are advantages over commercial LIMS solutions. In LIMS, modification of workflows requires in-depth knowledge of the data structure and the workflow tools and is thus restricted to trained personnel. Spontaneous changes of the scheme, to accommodate an observation in the experiment in the data collection, is thus more easily and rapidly done with the Phenotyper.

#### Mobile devices

As any standardised data acquisition system requires a higher 'activation energy' during implementation than the classical paper-based method, the attractiveness of the mobile device to the potential user is *a* if not *the* critical issue for the overall acceptance of the entire system. At the start of the project in 2011, we therefore recruited a beta-user group that was sociologically (age, computer-affinity, intellectual background) representative for the envisaged user population and had previous experience with phenotyping in greenhouse and fields. Brainstorming supplied the following criteria for high attractiveness of the device: large screens with good readability in sunlight, robust barcode-scanner, large, glove-compatible keyboard, and long battery lifetimes ( $\geq 8$  h). Additionally, low weight, easy handling of the device and a sun-light compatible barcode-scanner were asked for. Some of the criteria are obviously hard to combine, like low weight and handiness with large screens, bright displays and low weight with long battery lifetimes. We implemented the Phenotyper for barcode-scanner terminals (details see Material and methods) and provided configuration tools to adjust the GUI to different screen width and keyboard types of tablets and smartphones.

#### Graphical user interface on the mobile device

On-site, a GUI displays the user-defined phenotyping schemes for data entry into the personal digital assistant (PDA). A check function alerts if no scheme file is found or if a scheme file is invalid. The phenotyping scheme is selected on the start page of the GUI (Figure 3A). On this page, the identifier (id) of the object, on which the

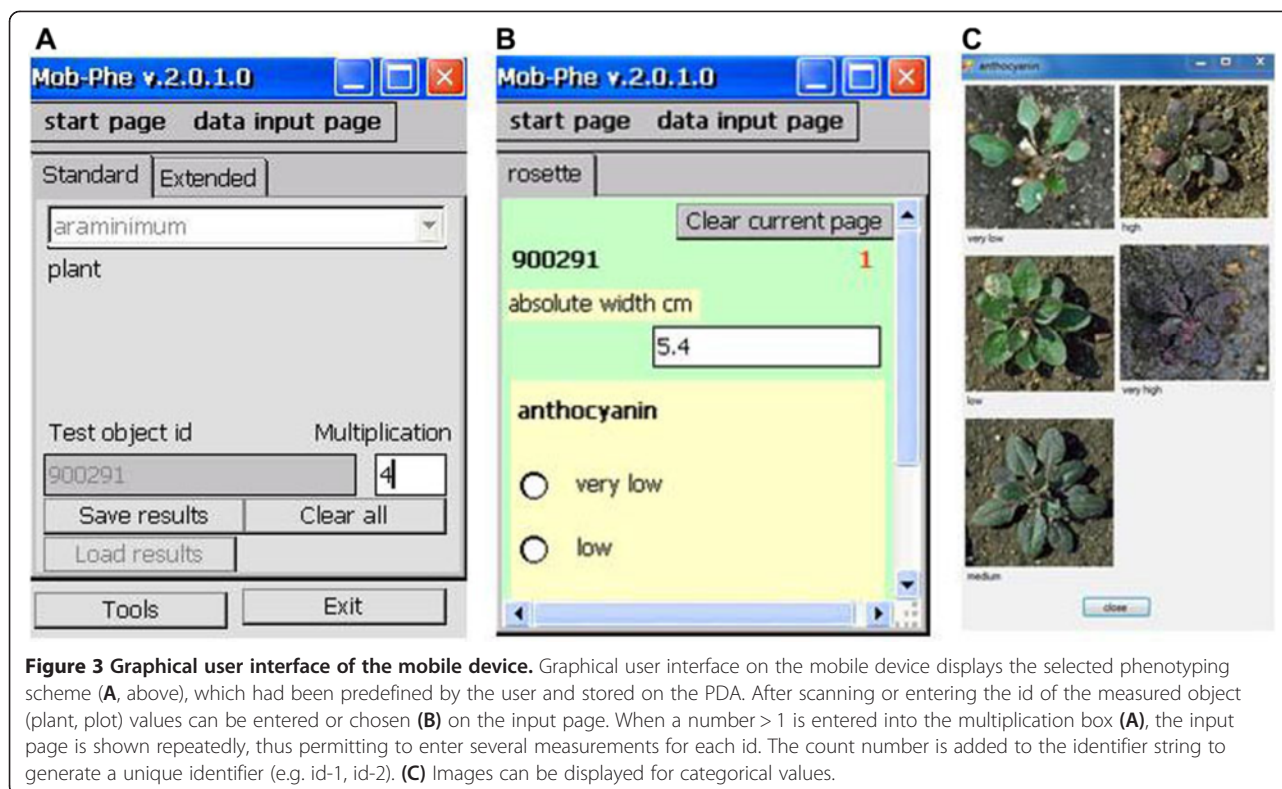
measurements are performed (e.g. a plant, a plot, a sample), is entered. The entities defined in the phenotyping scheme are displayed as subsequent pages, on which the parameters that are to be measured for each entity are listed (Figure 3B). Data can be entered (numeric variables, Figure 3B) or chosen (categorical variables, Figure 3B) from the listed attributes. Categorical variables can be illustrated by images (Figure 3C). A multiplication function is provided for scoring a single feature on a number of objects sharing one id, e.g. several plants in a plot. The entry of a multiplication factor on the first page defines the number of repeated measurements per id and permits efficient data entry for multiple measurements.

The Phenotyper generates unique names for the results files based on the device name, date and time information. With each measurement, date and time are recorded. To ensure the validity of these time stamps, the user is prompted to check the date and time settings and re-adjust it with a configuration tool. With the tool, the user can also define the storage location for phenotyping schemes and result data, assign functions (e.g. decimal separator, entry) to specific keys, and enable/disable the use of the end signals of barcode-scanners as entry signal. The tool thus facilitates adjusting the device to personal preferences without manual editing of the registry. When the factory default settings are restored by a hard reset, the tool automatically reactivates these personal settings, thus increasing the Phenotyper's reliability under harsh field conditions. Field tests furthermore revealed the need to check and edit data in the GUI. Individual entries can be selected and recalled to the GUI by choosing the id that acts as a primary key. The entry can then be modified and saved.

The optimization of the system for small screens makes the Phenotyper superior to LIMS systems, when data are to be recorded while moving through a greenhouse, field or natural eco-system. LIMS in general and the Phenopsis system [9,10], are optimized for automatic data import from machines (e.g. analytical devices, sensors) or for manual data entry on the LIMS user interface or on a web page, displayed on machines with large screens and network access. The Phenotyper can, however, transfer data directly from the mobile device to the LIMS Nautilus 9.0 (Thermo Fisher Scientific, Dreieich, Germany). We use this feature to record locations for plants and seed stocks [30]. However, as data storage *per se* does not require a sophisticated LIMS, we decided to store phenotypic data in an open-source database.

An efficient and meaningful use of the Phenotyper requires ids that are linked to information on the object, on which the phenotyping is performed. The optimal solution for the object identification is a unique id from a plant description table used by all users of the database, in which the phenotyping results are stored (R-DB) (see below). For





the TROST project (BMEL/FNR 22011208), that served as a test case, these ids were imported from the LIMS-based plant cultivation database of the MPI-MP [30] into the R-DB [31]. Alternative solutions for the documentation of plant cultivation experiments have been suggested [32-34]. When all plants of an institution or project are identified by a unique id, these soon have more than four digits making their entry tedious and error-prone, unless they are converted to barcodes. An alternative solution is the combination of experiment ids and unique ids within an experiment.

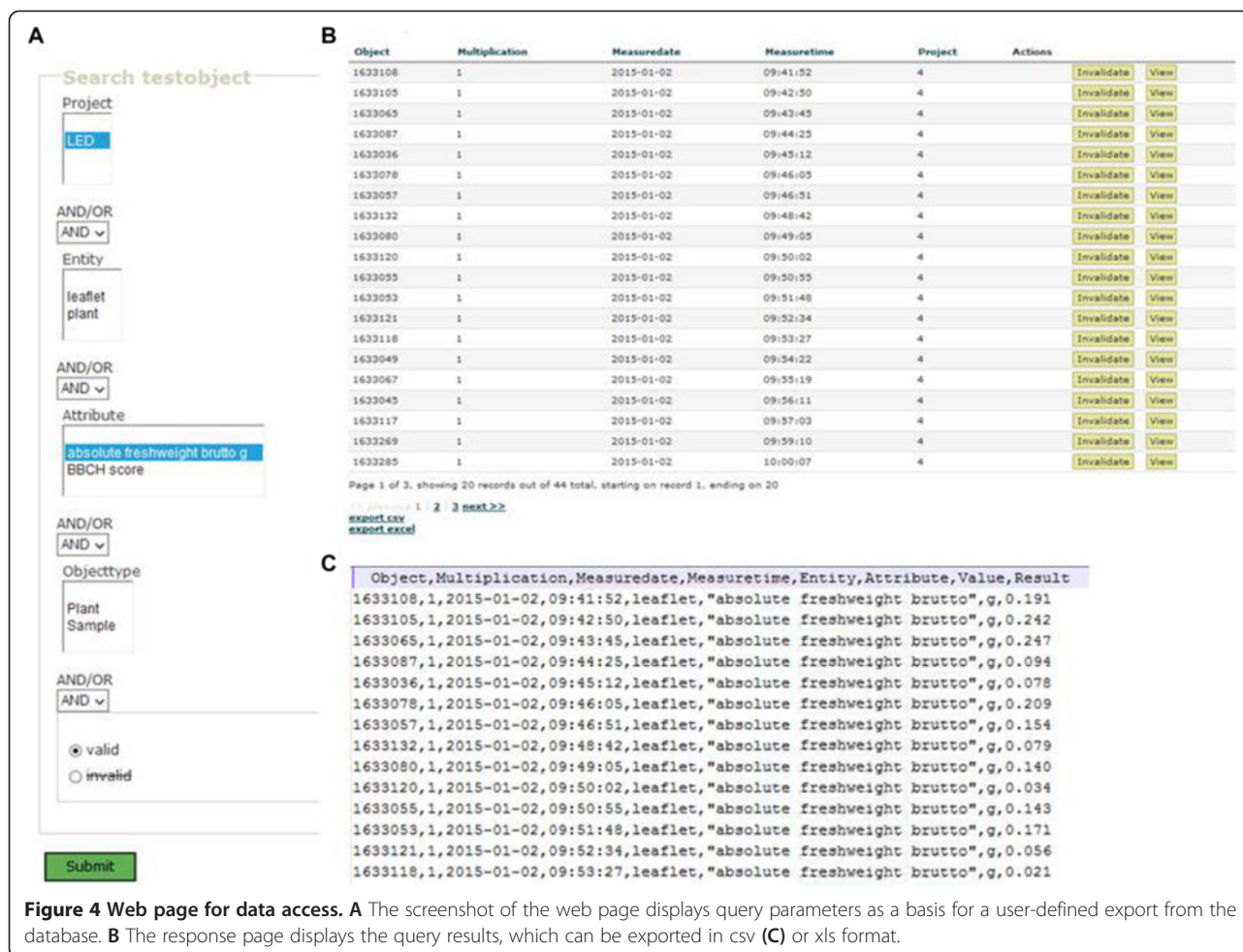
The Phen typer can also be used stand-alone without the link to a plant database containing meta-information. In this case, information about e.g. the genotype of the plant or the seed lot, the treatment, the cultivation location and the planting date could be recorded in a start scheme at the beginning of each experiment. The use of standard start schemes ensures that all relevant information is documented. Treatments and genotypes can be defined as attribute-value combination (e.g. genotype – Col-0, mutant – Wt, mutant – cbb). The efficiency of recording can be enhanced by grouping plants in experimental units that inherit information from the handling unit to its members [30]. In addition to enforcing metadata entry in standardised start schemes by institutional operation procedures, users can be motivated to enter metadata completely and correctly by the increased efficiency of data evaluation. Phen typer allows downloading all data

required for efficient data analysis in one dataset of highly standardised structure (see section data access).

#### Data transfer and quality control

On site, the GUI on the mobile device operates without access to WLAN or mobile phone networks, making the Phen typer usable in areas devoid of both (Faraday cage, remote areas). Data are stored immediately in the permanent memory of the mobile device to secure against data loss due to terminal crash, hard reset or power loss. The result data are stored in comma separated values format (csv) and XML format. In the csv-file (see Figure 4C), entries recorded for one identifier are recorded in one row. The cells' contents are identified by the headers in the first rows. Both formats can be read easily by spreadsheet programs like MS Excel. The csv format is suitable for import into statistics packages like SAS (SAS Institute Inc., Cary, NC, USA) or R [35]. When csv files from schemes containing different entity-attribute-value combinations are combined, the resulting file will contain empty cells for those variables that are not represented in all schemes.

If data from the csv-files are to be uploaded into the entity-attribute-value structure of a database, the file needs to be reformatted. Mapping-parsing scripts need to take into account that the order in which the variables are presented varies and that the array may contain empty cells. In XML files, however, every entry is tagged by the name of the variable. The upload procedure is

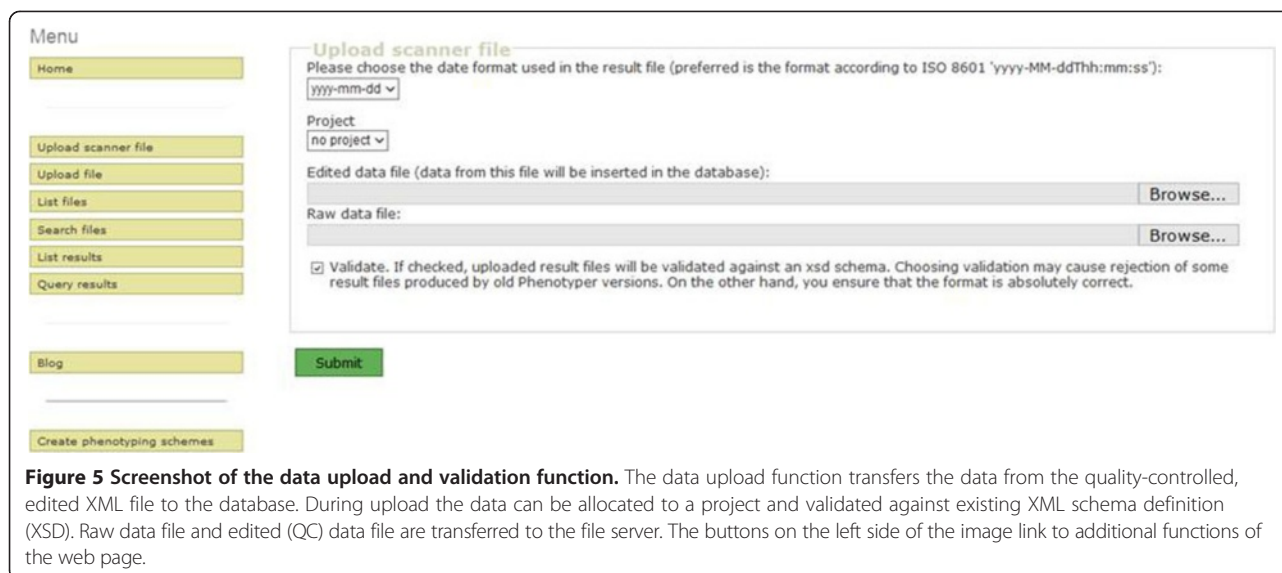


thus completely independent of the order, in which the variables are presented in the file, as long as the structure hierarchy is maintained. The number and order of variables per identifier can also vary within one file without causing empty cells (see Additional file 2). Furthermore, there are powerful validation tools to check XML files. Files in XML format can thus be easily used by automatic data processing tools, e.g. for the transfer of data into the database of a LIMS. The LIMS used at the MPI-MP, Nautilus 9.0 provides procedures for data import from csv and from xml files. We found the xml-import engine easier and safer to handle when data from one import file were to be written to several different LIMS tables.

For secure data storage, we developed a results database (see Data storage) combined with a file server for the storage of original (raw) and quality controlled results files. The quality control (QC) of the data for plausibility and typing errors is done after the retrieval of data from the mobile device and before upload. Immediate QC is a most important step to ensure high data quality, as it allows re-evaluation of out of range results. However, modification of the original dataset may lead to accusation of

fraudulent data modification. The original file from the scanner should therefore be stored together with the file resulting from QC to safeguard the user and establish good data management practise. To our knowledge, the Phenotyper is the first system that provides this feature.

For data transfer from the mobile device or local computer disk to the database server, an upload function is provided on a web page (Figure 5). The web page performs a quality control on the file structure, copies the files to the file storage system, uploads the data from the QC file to the results database and links the imported datasets to the files. The name of the user who logged onto the transfer web page is added to the uploaded datasets as an ownership-stamp together with the upload time stamp. This information forms the basis for the maintenance of intellectual property rights and the access policy (see section Data access). To facilitate data sharing in projects, the user can choose from the list of projects, to which he is affiliated, which project the uploaded dataset belongs to. The default setting is 'no project' for data that are not meant to be shared.



### Data storage

Result data are stored in a results database (R-DB) modified from a prototype developed for the TROST project [31]. In the relational R-DB, data are stored in an entity-attribute-value (EAV) data model (ER diagram see Additional file 1: Figure S2). This data storage structure is especially suitable for so-called sparse and volatile data, which means that the number of parameters measured on an object is small compared to the number of parameters that could be measured and it tends to change during a project [31,36]. As an example, at the beginning of a project, 20 parameters may be assessed on each plant. After the first evaluations, measurements are continued only on the five most predictive parameters plus a newly discovered parameter. In classical data storage structure, this addition would require a change in the database structure and result in many empty cells. In the EAV system, the identifier will be combined with attribute-value combinations for 20 parameters at the beginning of the project and change to six identifier-parameter combinations in a later phase. The query, with which data for a parameter of interest are extracted, remains the same as the automatic data evaluation procedures throughout the project.

Several other projects [24,37,38] and the commercial system Lemnatec (Aachen, Germany) use relational (SQL) databases for raw data storage, whereas the image storage and processing system IAP uses the no-SQL database MongoDB [8,25].

The user data in the R-DB are the basis for the access regulation to the upload web page, the ownership stamp, the project information and the data access during download. The link between projects and user names is modelled in a m - n relationship, thus each user can belong to several projects and each project can be shared by several users.

To permit data interpretation independent of an access to the S-DB, controlled vocabulary needs to be imported into the R-DB. In our implementation, controlled vocabulary was mirrored from the S-DB to the R-DB. An alternative solution is to import the controlled vocabulary together with its primary key from the results files to the R-DB. In our implementation, the R-DB should only be connected to one S-DB to avoid ambiguity. One S-DB can, however, serve many R-DBs. The R-DB could either be hosted locally by an individual researcher or research group or centrally by an institute or project coordination site.

To facilitate the process of setting up web server and R-DB locally, we develop a setup routine for both (see Additional files 3, 4 and 5). The person or institution hosting the Phenotyper would, however, need to have some experience with web server and database administration; thus, some degree of centralization in an institution or larger project is desirable. We do not, however, envisage a central R-DB maintained for many research projects for a decade or longer. As many phenotyping projects will be financed by time-limited grants, interest in and maintenance off the central R-DB may wane at some stage. When checking the availability of central data storage solutions, we found that some of them were no longer accessible or maintained less than ten years after publication [22,39]. Institutions that concentrate on scientific discovery usually have no resources to maintain databases for decades when these are no longer actively used. Thus, data ought to be transferred to a central long-term storage facility that is dedicated to provide quality-controlled data storage for many decades. Examples for centrally maintained long-term storage facilities are the EWIG project, the RADAR and the DRYAD facility [40-42]. These archives provide a document identifier and meta-information for each



dataset and ensure controlled access to the data, but not necessarily a fast data transfer structure. The DataCite metadata store provides metadata and document identifiers for dataset that are stored in different archives and thus allows efficient search for datasets and their citation [43]. The best data storage format is an ASCII or XML format as these formats can be read independently of specific software products. Our R-DB provides a function to generate standardised ASCII report, in which each dataset contains the complete information in a structured format. These data can later on be easily imported and evaluated by other scientists, then increasing the awareness of the work and of the person, who deposited the data [24].

A trigger for the transfer of raw phenotyping data to the long-term storage facility could be the publication of results based on these data. This procedure is already state of the art in research areas like transcriptomics [44] or in large ecological studies [42]. Universities could make the deposition of all raw data, on which a PhD thesis is based, an obligatory part of the defence. Furthermore, data should be committed to long-term storage at the end of a project to maintain long-term data access as demanded by an increasing number of funding agencies (e.g. Deutsche Forschungsgemeinschaft [45], EU, BBSRC [46]).

#### Data access

For data evaluation or submission to long-term storage facilities, data need to be downloaded from the database. To provide easy access to a potentially large number of datasets, we generated a web page with search and export functions (Figure 4A). The search function provides filters for entries e.g. for entities and attributes to limit the export to data required for an evaluation. The data are displayed on the web page (Figure 4B) and exported in csv or xls format (Figure 4C). Data access is regulated based on the ownership stamp, project information, the upload time stamp and the embargo period defined for the project or institution. By default, data are only accessible to the user defined in the owner field. Data assigned to a project are accessible to all users affiliated to the project. The timestamp opens data access to all users after the end of predefined embargo period.

Whether the ownership stamp is sufficient to guard intellectual property rights is debatable. The ownership stamp links to a table containing first and family name of the user, which may be sufficient to identify a person within an institution when combined with information recorded in the Human Resources Department of the respective institution. A potential alternative could be the inclusion of the scientist's Scopus identifier into the user management.

Data access to centrally stored result data is a sensitive issue. In contrast to the situation in material sciences [24], many scientists in life sciences are very reluctant to transfer their raw data to a central depository for fear of

uncontrolled data access and use by other scientists. This lack of enthusiasm is enhanced by the knowledge that raw data are not covered per se by the regulations on intellectual property rights. Thus, in addition to providing a system that facilitates structured recording and storage of data and to provide long-term storage facilities, some issues linked to open data access remain to be solved by discussion in the scientific community. One option is to agree on making it obligatory to provide raw data on publication, when submitting a thesis and at the end of a project. Furthermore, raw datasets with metadata ought to be made searchable and citable as suggested by the DataCite initiative, thus conferring value to the dataset *per se* and rewarding data sharing.

#### Conclusion

Altogether, our phenotyping tool provides a well-structured, but flexible data acquisition and management structure for on-site measurements with mobile devices like smartphones or barcode-scanner terminals. As the user chooses the type of controlled vocabulary, the tool can be used for any type of manual data entry by any discipline, phenotyping as well as documentation of field sampling to inventories in a (museum) collection. We have used the Phenotyper for all these purposes successfully. Likewise, measurement data can be stored alongside genotype or treatment data including spatial or experimental design information. The predefined structure and the use of controlled vocabulary facilitate efficient automatic data evaluation and shared data use. Mechanisms for the protection of intellectual property rights and making datasets citable per se increase the scientist's trust when submitting his data to a central depository. Upon publication, well-structured files with controlled vocabulary can be automatically generated from the database, which saves much time and avoids errors compared to manually compiling a data file from many sources (e.g. spreadsheet files). In contrast to self-defined parameter, no definition needs to be written for controlled vocabulary. The submission of raw data to long-term storage facilities is thus made very simple and will guarantee data accessibility beyond the live time of the results database of an individual project. Thus, the tool may help to improve data management within phenotyping projects and, by automatic generation of structured csv export files, long-term data accessibility.

#### Material and methods

##### Software and database

The web pages for combining controlled vocabulary to phenotyping schemes and for the administration of the S-DB was implemented with Framework Struts 2 (Apache Software Foundation) within the application development system MyEclipse (Genuitec, Flower Mound, TX 75028, US). The servlet container was Apache Tomcat 6.0



(Apache Software Foundation). The web pages for the upload and download of result data were implemented in Framework Cake-PHP 1.3 (Cake Software Foundation). The support web pages and the fileserver system were based on the system developed for the TROST project [31]. The databases S-DB and R-DB were implemented in MySQL 5.5 (Oracle, Redwood City, US); MySQL-Workbench 5.2 (Oracle) was used for administration. The mirror function that imports user data and controlled vocabulary from the S-DB to the R-DB was implemented in SQL.

The graphical user interface for the mobile devices was implemented in .NET-Compact-Framework 3.5 (Microsoft Corporation) in C#. The application development system was Microsoft Visual Studio 2008 (Microsoft Corporation).

The internationalization was realised in frameworks Struts 2, Cake-PHP and the .NET library.

#### Hardware

The Phenotyper can be operated on mobile devices, e.g. tablets, barcode-scanner terminals or mobile phones with the operating systems Windows 7 Pro, Windows Mobile 6.0, Windows CE 5.0. The use on other Windows based systems is possible but was not tested extensively. During implementation and beta-testing, we used the barcode-scanner terminals Datalogic elf 00ALOLS-1 N1-Meno with laser-scanner (Datalogic, Diez, Germany) and M3T-1D (M3-Mobile, Seoul, Korea) and the semi-robust tablet computer Advantech-DLoG PWS 770 (ADVANTECH, München, Germany).

#### Availability

Users who like to see whether the functionality of the systems fits their requirements without having to run a complete setup can obtain test accounts for the scheme composer web pages and a test version of the results database. Login information is available on demand from the authors for a limited time period.

The complete software package is available as Additional files 3, 4 and 5 submitted with the manuscript. These files contain the installation package for the mobile device GUI, the files for the web pages and dumps of the databases. The software (see Material and methods) required to run e.g. the web pages is open-source. Links to repositories, from which the open-source software can be downloaded are provided in the help files of each package.

The software package and the source files required for further development of the software by the community is available from the software project repository of the Bioinformatics organization ([http://www.bioinformatics.org/groups/?group\\_id=1210](http://www.bioinformatics.org/groups/?group_id=1210) and <http://www.bioinformatics.org/cgi-bin/viewvc.cgi/phenotyper/> for the source code).

#### Additional files

**Additional file 1: ER diagrams.** Entity-relationship diagrams for the S-DB (Figure S1) and the R-DB (Figure S2).

**Additional file 2: Example file.** Example of data structures of information collected on one object on several different dates. **Table A1.** shows the data the csv format, in which they are exported from the database, **Table A2.** shows the data after reformatting in the 'broad' format, in which data are generally presented in spreadsheet programs, **Table A1.** shows the data in xml format.

**Additional file 3: PDA-Software.** This archive contains the software to be installed on the personal digital assistant (PDA) including configuration files and installation instructions.

**Additional file 4: Phenotyping Scheme Composer.** This archive contains a dump file for the S-DB, a WAR file for installing the web application for generating phenotyping schemes, an encrypting tool, and installation instructions.

**Additional file 5: Phenotyper web site and results database.** This archive contains a dump file for the R-DB, the root directory of the result web application, an encrypting tool, and installation instructions.

#### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

JG performed the implementation, selected the hardware and contributed to the manuscript, KK designed the system, organised the field tests and wrote the manuscript. Both authors tested the system. Both authors read and approved the final manuscript.

#### Acknowledgements

The German Research Foundation funded the work by grant DFG KO1731/3-1 to Karin Köhl. The following people were part of the beta-test group and contributed, by their suggestions and feedbacks, to the optimization of the system: Jessica Alpers, Ulrike Seider, Heike Sprenger and Ellen Zuther (MPI-MP), Katharina Rudack, Sylvia Seddig and Regina Schmidt (JKI Groß Lüsewitz), Vanessa Prigge (SAKA, Windeby), B. Truberg (Norika, Groß Lüsewitz). We thank Axel Fischer (MPI-MP) for valuable comments on the manuscript and Kenny Billiau for assistance with the implementation of the web pages.

Received: 12 January 2015 Accepted: 19 March 2015

Published online: 07 April 2015

#### References

1. Tester M, Langridge P. Breeding technologies to increase crop production in a changing world. *Science*. 2010;327(5967):818–22.
2. Finkel E. With 'Phenomics', plant scientists hope to shift breeding into overdrive. *Science*. 2009;325(5939):380–1.
3. Furbank RT, Tester M. Phenomics - technologies to relieve the phenotyping bottleneck. *Trend Plant Sci*. 2011;16(12):635–44.
4. Houle D, Govindaraju DR, Omholt S. Phenomics: the next challenge. *Nat Rev Genet*. 2010;11(12):855–66.
5. Traitmill -Platform and Process. [[http://www.cropdesign.com/tech\\_traitmill.php](http://www.cropdesign.com/tech_traitmill.php)]
6. The Plant Accelerator. [<http://www.plantaccelerator.org.au/>]
7. JPPC - the Jülich Plant Phenotyping Centre. [[http://www.fz-juelich.de/ibg/ibg-2/DE/Organisation/JPPC/JPPC\\_node.html](http://www.fz-juelich.de/ibg/ibg-2/DE/Organisation/JPPC/JPPC_node.html)]
8. Klukas C, Chen D, Pape J-M. Integrated analysis platform: an open-source information system for high-throughput plant phenotyping. *Plant Phys*. 2014;165(2):506–18.
9. Granier C, Aguirrezabal L, Chenu K, Cookson S, Dauzat M, Hamard P, et al. PHENOPSIS, an automated platform for reproducible phenotyping of plant responses to soil water deficit in *Arabidopsis thaliana* permitted the identification of an accession with low sensitivity to soil water deficit. *New Phytol*. 2006;16:623–35.
10. Fabre J, Dauzat M, Negre V, Wuyts N, Tireau A, Gennari E, et al. PHENOPSIS DB: an Information System for *Arabidopsis thaliana* phenotypic data in an environmental context. *BMC Plant Biol*. 2011;11(1):77.
11. Cobb JN, DeClerck G, Greenberg A, Clark R, McCouch S. Next-generation phenotyping: requirements and strategies for enhancing our understanding

- of genotype-phenotype relationships and its relevance to crop improvement. *Theor Appl Genet.* 2013;126(4):867–87.
12. Araus JL, Cairns JE. Field high-throughput phenotyping: the new crop breeding frontier. *Trend Plant Sci.* 2014;19(1):52–61.
  13. Dhondt S, Wuyts N, Inze D. Cell to whole-plant phenotyping: the best is yet to come. *Trend Plant Sci.* 2013;18(8):433–44.
  14. Zeeberg B, Riss J, Kane D, Bussey K, Uchio E, Linehan WM, et al. Mistaken identifiers: gene name errors can be introduced inadvertently when using Excel in bioinformatics. *BMC Bioinformatics.* 2004;5(1):80.
  15. Lancashire PD, Bleiholder H, Van Den Boom T, Landgeludde P, Strauss R, Weber E, et al. A uniform decimal code for growth stages of crops and weeds. *Ann Appl Biol.* 1991;119(3):561–601.
  16. Institute IRR. Standard evaluation system for rice. Los Banos, Laguna, Philippines: International Rice Research Institute; 1996.
  17. Baxter I, Ouzzani M, Orcun S, Kennedy B, Jandhyala S, Salt D. Purdue ionomics information management system. An integrated functional genomics platform. *Plant Physiol.* 2007;143:600–11.
  18. van Ommen B, Bouwman J, Dragsted L, Drevon C, Elliott R, de Groot P, et al. Challenges of molecular nutrition research 6: the nutritional phenotype database to store, share and evaluate nutritional systems biology studies. *Genes Nutr.* 2010;5(3):189–203.
  19. Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, et al. NCBI GEO: mining tens of millions of expression profiles—database and tools update. *Nucleic Acids Res.* 2007;35 suppl 1:D760–5.
  20. Goff SA, Vaughn M, McKay S, Lyons E, Stapleton AE, Gessler D, et al. The iPlant collaborative: cyberinfrastructure for plant biology. *Front Plant Sci.* 2011;2:34.
  21. Li Y-F, Kennedy G, Ngoran F, Wu P, Hunter J. An ontology-centric architecture for extensible scientific data management systems. *Int J Grid Comput Escience.* 2013;29(2):641–53.
  22. Podd Data Repository Project to aid Phenomics. [<http://www.plantphenomics.org.au/projects/podd/>]
  23. Phenoscape. [<http://kb.phenoscape.org>]
  24. Yang E, Matthews B, Wilson M. Enhancing the core scientific metadata model to incorporate derived data. *Future Generat Comput Syst.* 2013;29(2):612–23.
  25. Klukas C, Pape J-M, Entzian A. Analysis of high-throughput plant image data with the information system IAP. *J Integr Bioinform.* 2012;9(2):191.
  26. Slovak R, Goeschl C, Su X, Shimotani K, Shiina T, Busch W. A scalable open-source pipeline for large-scale root phenotyping of *Arabidopsis*. *Plant Cell.* 2014;26(6):2390–403.
  27. Triticeae Toolbox. [<http://triticeaetoolbox.org>]
  28. DbNP. [<http://www.dbnp.org>]
  29. Jaiswal P, Avraham S, Ilic K, Kellogg EA, McCouch S, Pujar A, et al. Plant Ontology (PO): a controlled vocabulary of plant structures and growth stages. *Comp Funct Genom.* 2005;6(7–8):388–97.
  30. Köhl KI, Basler G, Luedemann A, Selbig J, Walther D. A plant resource and experiment management system based on the Golm Plant Database as a basic tool for omics research. *Plant Methods.* 2008;4:11.
  31. Billiau K, Sprenger H, Schudoma C, Walther D, Köhl KI. Data management pipeline for plant phenotyping in a multisite project. *Func Plant Biol.* 2012;39(11):948–57.
  32. Hannemann J, Poorter H, Usadel B, Blasing OE, Finck A, Tardieu F, et al. Xeml Lab: a tool that supports the design of experiments at a graphical interface and generates computer-readable metadata files, which capture information about genotypes, growth conditions, environmental perturbations and sampling strategy. *Plant Cell Environ.* 2009;32(9):1185–200.
  33. Nieuwland J, Sornay E, Marchbank A, de Graaf B, Murray J. Phytotracker, an information management system for easy recording and tracking of plants, seeds and plasmids. *Plant Methods.* 2012;8(1):43.
  34. Exner V, Hirsch-Hoffmann M, Gruissem W, Hennig L. PlantDB - a versatile database for managing plant research, vol. 4. 2008. p. 1.
  35. R-Forge. [<http://r-forge.r-project.org>]
  36. Dinu V, Nadkarni P. Guidelines for the effective use of entity-attribute-value modeling for biomedical databases. *Int J Med Inform.* 2007;76(11–12):769–79.
  37. Mabee P, Balhoff J, Dahdul W, Lapp H, Midford P, Vision T, et al. 500,000 fish phenotypes: the new informatics landscape for evolutionary and developmental biology of the vertebrate skeleton. *J Appl Ichthyol.* 2012;28:300–5.
  38. Klukas C, Schreiber F. Integration of -omics data and networks for biomedical research with VANTED. *J Integr Bioinform.* 2010;7(2):112.
  39. Chern C-G, Fan M-J, Yu S-M, Hour A-L, Lu P-C, Lin Y-C, et al. A rice phenomics study-phenotype scoring and seed propagation of a T-DNA insertion-induced rice mutant population. *Plant Mol Biol.* 2007;65(4):427–38.
  40. EWIG. [<http://ewig.gfz-potsdam.de>].
  41. RADAR. [<http://www.radar-projekt.org/>].
  42. DRYAD. [<http://datadryad.org/>].
  43. DataCite. [<https://www.datacite.org>].
  44. GEO. [<http://www.ncbi.nlm.nih.gov/geo/>].
  45. Empfehlungen 200901. [[http://www.dfg.de/download/pdf/foerderung/programme/lis/ua\\_inf\\_empfehlungen\\_200901.pdf](http://www.dfg.de/download/pdf/foerderung/programme/lis/ua_inf_empfehlungen_200901.pdf)].
  46. Data sharing policy. [<http://www.bbsrc.ac.uk/web/FILES/Policies/data-sharing-policy.pdf>].

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

