Check for updates

SOFTWARE TOOL ARTICLE

REVISED PhenoApp: A mobile tool for plant phenotyping to record field and greenhouse observations [version 2; peer review: 2 approved]

Franco Röckel [ID]1*, Toni Schreiber2*, Danuta Schüler3*, Ulrike Braun1, Ina Krukenberg4, Florian Schwander [ID]1, Andreas Peil5, Christine Brandt [ID]6, Evelin Willner7, Daniel Gransow7, Uwe Scholz [ID]3, Steffen Kecke2, Erika Maul1, Matthias Lange [ID]3, Reinhard Töpfer1

1Julius Kühn Institute (JKI) - Federal Research Centre for Cultivated Plants, Institute for Grapevine Breeding Geilweilerhof, Siebeldingen, 76833, Germany
2Julius Kühn Institute (JKI) - Federal Research Centre for Cultivated Plants, Data Processing Department, Erwin-Baur-Straße 27, Quedlinburg, 06484, Germany
3Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Corrensstraße 3, Seeland, 06466, Germany
4Julius Kühn Institute (JKI) - Federal Research Centre for Cultivated Plants, Data Processing Department, Königin-Luise-Strasse 19, Berlin, 14195, Germany
5Julius Kühn Institute (JKI) - Federal Research Centre for Cultivated Plants, Institute for Breeding Research on Fruit Crops, Pillnitzer Platz 3a, Dresden/Pillnitz, 01326, Germany
6Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), The Satellite Collections North, Parkweg 3a, Sanitz, 18190, Germany
7Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), The Satellite Collections North, Inselstraße 9, Malchow/Poel, 23999, Germany

* Equal contributors

Abstract

With the ongoing cost decrease of genotyping and sequencing technologies, accurate and fast phenotyping remains the bottleneck in the utilizing of plant genetic resources for breeding and breeding research. Although cost-efficient high-throughput phenotyping platforms are emerging for specific traits and/or species, manual phenotyping is still widely used and is a time- and money-consuming step. Approaches that improve data recording, processing or handling are pivotal steps towards the efficient use of genetic resources and are demanded by the research community. Therefore, we developed PhenoApp, an open-source Android app for tablets and smartphones to facilitate the digital recording of phenotypical data in the field and in greenhouses. It is a versatile tool that offers the possibility to fully customize the descriptors/scales for any possible scenario, also in accordance with international information standards such as MIAPPE

Open Peer Review

Approval Status ✓ ✓

	1	2
version 2 (revision) 28 Nov 2022	✓ view	✓ view
	↑	↑
version 1 07 Jan 2022	? view	? view

1. **Karin Köhl** [ID], Max Planck Institute of Molecular Plant Physiology, Potsdam, Germany

2. **Andrew P. French**, University of Nottingham,

(Minimum Information About a Plant Phenotyping Experiment) and FAIR (Findable, Accessible, Interoperable, and Reusable) data principles. Furthermore, PhenoApp enables the use of pre-integrated ready-to-use BBCH (Biologische Bundesanstalt für Land- und Forstwirtschaft, Bundessortenamt und CHemische Industrie) scales for apple, cereals, grapevine, maize, potato, rapeseed and rice. Additional BBCH scales can easily be added. The simple and adaptable structure of input and output files enables an easy data handling by either spreadsheet software or even the integration in the workflow of laboratory information management systems (LIMS). PhenoApp is therefore a decisive contribution to increase efficiency of digital data acquisition in genebank management but also contributes to breeding and breeding research by accelerating the labour intensive and time-consuming acquisition of phenotyping data.

Keywords
Android app, plant phenotyping, digital data acquisition, LIMS, BBCH, FAIR principles

Nottingham, UK

Any reports and responses or comments on the article can be found at the end of the article.

This article is included in the Plant Science gateway.

Corresponding author: Franco Röckel (franco.roeckel@julius-kuehn.de)

Author roles: Röckel F: Conceptualization, Software, Writing – Original Draft Preparation; **Schreiber T**: Software, Writing – Original Draft Preparation; **Schüler D**: Software, Writing – Original Draft Preparation; **Braun U**: Software; **Krukenberg I**: Software; **Schwander F**: Funding Acquisition, Supervision, Writing – Review & Editing; **Peil A**: Conceptualization, Software; **Brandt C**: Data Curation, Validation; **Willner E**: Data Curation, Validation; **Gransow D**: Data Curation, Software, Validation; **Scholz U**: Funding Acquisition; **Kecke S**: Supervision; **Maul E**: Funding Acquisition, Supervision, Writing – Review & Editing; **Lange M**: Funding Acquisition, Writing – Original Draft Preparation; **Töpfer R**: Funding Acquisition, Supervision, Writing – Review & Editing

> **REVISED** **Amendments from Version 1**
>
> This new version of the article does not include any major changes in the shape and contents of the original one. However, some minor changes to the wording of the text have been made in the methods section to meet clarifications of points raised by the peer-reviewers. A reference for the LIMSOPHY (laboratory information management system) and the use case section has been added.
>
> **Any further responses from the reviewers can be found at the end of the article**

Introduction

Classical breeding of agronomical improved crops relies on crossing of desired genotypes, growing their offspring and performing genotypic and phenotypic selection on the various traits. Desired genotypes consist of elite breeding material as well as genetic resources that are characterized. Approximately 7.4 million accessions are stored in around 1750 *ex-situ* germplasm collections (genebanks) worldwide.[1] These genebanks contain genetically and phenotypically diverse plant material and are excellent resources of novel traits useful for future plant breeding purposes in the context of changing demands.

Although innovative breeding technologies like marker-assisted selection have dramatically evolved to improve selection accuracy and intensity during recent decades,[2,3] only a small number of genotypes contributed directly to modern crop cultivars mainly due to the lack of sufficient phenotypic and genotypic characterization or limited evaluation of agronomic traits.[4–6] The efficient use of germplasm collections as a source of genetic variation is time consuming and arduous and therefore still remains a challenging task. Thus, with the development and decreasing costs of genotyping and sequencing technologies, genebank phenomics display the current bottleneck for the utilization of genetic resources.[7–9] Nevertheless, tremendous advancements in high-throughput phenotyping technologies during recent years are closing this gap.[10–12] However, as long as these high-throughput phenotyping technologies are a limiting factor, e.g. due to missing technologies, infrastructure or funding, supporting tools for manual data acquisition and recording can accelerate and support the evaluation of genetic resources and can increase breeding efficiency.

Plant phenomics is a multidisciplinary field developing novel sensing and imaging techniques for high-throughput phenotyping of plant genetic resources, and has applications in breeding.[13–15] The research basis are morphological, agronomical, physiological and metabolic features, whereas the handling, processing and adequate utilization of data is still challenging. The big advantage of high-throughput platforms compared to manual plant phenotyping is its objectivity as well as the time and cost-effectiveness. However, the development is slow and a plethora of research needs to be done to provide further valuable screening tools with practical use in the future.[16] With all these developments, the volume of data potentially usable for plant breeding has increased rapidly and will further increase in future. Therefore, the type of data generated range not only from agronomic and breeding-relevant phenotypic data, to results from quantitative and qualitative genetics, but can contain further information on fertilization, plant protection, field and soil conditions, geodata and weather data. Most data sets differ not only in their object of investigation, but also in their type, format and context of origin. The establishment of the FAIR principles (Findable, Accessible, Interoperable and Reusable) are therefore an important basis regarding harmonization and can increase the efficiency of plant breeding.[17–19] Another requirement is the interoperability of data in terms of machine-readable access to support continuous data analysis flows. To address the resulting challenges, internationally recognized work has already been done in the area of metadata and the minimum information standard for plant phenotyping data - MIAPPE - has been developed.[20] In addition, a large number of semantic resources exist, e.g. AGROVOC (word combination of agriculture and vocabulary)[21] or the general ontologies of the Open Biological and Biomedical Ontology (OBO) Foundry.[22] In the area of data structures, concepts for generalization have been proposed, such as Investigation-Study-Assay (ISA-TAB)[23] or, more recently, the Core Scientific Dataset Model,[24] which abstracts individual data structures to a self-describing generic data structure.

To ensure proper understanding of the content, we rely on common assessment and observation protocols, vocabularies and units. Many initiatives use their dedicated scoring scheme or, in the case of collaborative efforts such as genebanks, a commonly agreed set of observation protocols.[25] Prominent examples for scoring schemes are Darwin Core (DwC),[26] BBCH (Biologische Bundesanstalt für Land- und Forstwirtschaft, Bundessortenamt und CHemische Industrie) for phenology[27] and the International Union for the Protection of New Varieties of Plants (UPOV) lists for morphological traits. The manual acquisition of such field phenotyping data into local infrastructures is a first component in a data publication pipeline.

A tool support for field phenotyping, i.e. observation scoring, enables the re-usability and fulfils the FAIR criteria. For this purpose, genebanks, research units, breeding companies or other stakeholder in plant phenotyping use not only simple, handwritten records, but also simple digital forms of recording. Microsoft Excel installations on mobile devices with or without voice input support, and solutions closely embedded in individual database infrastructures, such as the Genebank Information System/Bonitur (GBIS/B) at the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK)[28] or recently published apps like vitisBerry,[29] SeedCounter[30] or Plant Screen mobile[31] may address specific demands, but lack convenience or generalization. As an alternative to such proprietary, hand crafted and/or non-generic solutions for recording and transferring field-scoring data into a digital representation or even databases such as Lab Information Management Systems (LIMS), an in-field digitalization of observations on a general basis represents the initial process to advance the field-data acquisition process.

In the current manuscript, we present an easy-to-use Android app (PhenoApp) for recording field and greenhouse observations on mobile devices to relieve the labor intensive and time-consuming process of manual phenotyping. All evaluated descriptors/scales can be individually created with maximum customization, allowing a fast and seamless data transfer into a digital representation for further utilization in genebank management and/or breeding research.

Methods
Implementation
Data import and export

After installation, the main directory contains an 'in' and 'out' folder. Input files need to be copied in the 'in' folder. Sample input files are then automatically displayed and selectable in the app entry page. The save button in the upper right corner of the app (Figure 1) creates an output Excel file saved in the 'out' folder.

Input file with descriptor list

The first sheet of the input file (Locations) contains information on the specific plant locations with the columns plot, row, plant, accession name and number, variety number, genotype, mother, father, information and database-key. The last column (database-key) is not visible in the application and is used in the output file for better synchronization with the database. Only the first three columns or as alternative the fourth column with information on the exact plant locations are obligatory fields, all others are optional and implemented for better data handling and management as well as to give the user additional orientation and information.
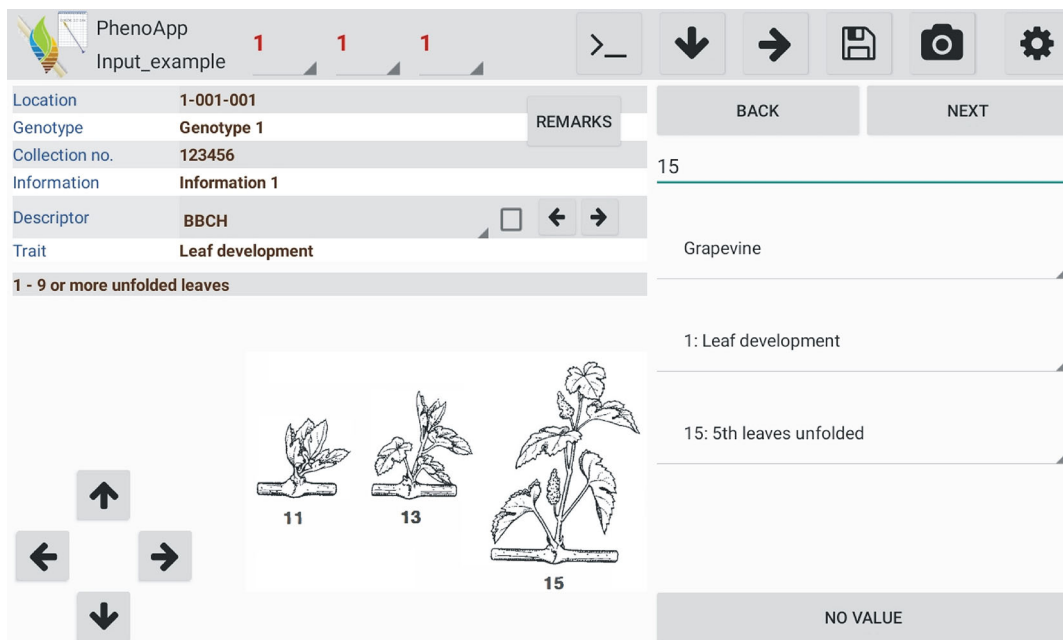

Figure 1. General app view with a data-recording example for a BBCH (Biologische Bundesanstalt für Land- und Forstwirtschaft, Bundessortenamt und CHemische Industrie) growth stage of grapevine.

The second sheet (Traits) contains all descriptors. These are described in the five columns which are shortcut (trait abbreviation), description, type, values and remark. The column type needs to be filled with 1 to 5 according to the format type of the recorded phenotypic data:

1. Rating: For the acquisition of categorical data. The scale needs to be specified in the value column (for example, 1: none; 3: little; 5: medium; 7: strong; 9: very strong).

2. Measured value: PhenoApp displays a numeric keypad for data entry. Units can be specified in the remarks field of the individual trait.

3. Date: PhenoApp displays selectable buttons for today, yesterday, day before yesterday, tomorrow, day after tomorrow and a button for an individual date selection. Data are stored in numerical date format.

4. Text: PhenoApp opens a keyboard for free text input.

5. BBCH: PhenoApp displays the pre-integrated BBCH scales for phenotyping of phenology.

An example input file is provided in *Underlying data.*[35]

Output file

The structure and content of the output file is the same as for the first sheet of the input file (specific plant locations) with additional columns for remarks and phenotyped data. Output format can be an Excel file or a csv (comma separated values) file for better machine readability. An example output file is provided in *Underlying data.*[35]

Important app functions

Phenotyping phenology based on BBCH scales

As an additional mode for a convenient phenotyping of phenology over a certain time, PhenoApp provides pre-integrated BBCH scales without the need of own descriptors. Furthermore, if the option "BBCH question" is selected, each time PhenoApp is opened, it asks if the creation of a BBCH entry with the current date is desired. The integrated scales contain three list levels from species to principal growth stage to specific growth stage (Figure 1). For comfort during phenotyping, the list levels species and principal growth stage will be copied so that only the specific growth stage needs to be selected at every location entry.

The release version has the BBCH scales of apple, cereals, grapevine, maize, potato, rapeseed and rice already implemented.

If needed, the user can implement scales of other crops or add own scales using an Excel template (bbch_template.xls) located in the 'bbch' folder of the app main directory (see *Software availability*[36]). The first sheet is for the species as well as the principal growth stages and all further sheets for the specific growth stages. In addition, there is the option of adding one sample image for each principal growth stage, which can be displayed in the app. To do this, the images need to be stored with the same filename as the stages in the app's 'bbch' folder together with the template file. The next time PhenoApp is started, the new BBCH scale will be imported to the internal database for further use and files are automatically deleted from the 'bbch' folder.

Displaying sample images of descriptors

If a sample image of a descriptor with the same name as the descriptor shortcut are copied in the 'descriptorPictures' folder of the main directory, PhenoApp displays that image during phenotyping the same way as for BBCH scales.

Taking pictures with the device camera

The user can take photos without restrictions during application. Clicking on the camera icon automatically opens the device standard photo app. Once a desired photo has been taken, it is saved in the subfolder 'fotos' in the app's main

directory. The file name of the photo is a combination of the currently specified location, the actual time stamp and the variety name (if given).

Taking general and location specific notes

Centered in the upper half of the screen is a 'REMARKS' button, which opens a window with two text fields for the input of specific location information and any additional information on the current plant, respectively. The text entries of both fields are included in the output file in separated columns. This feature gives the user the possibility to take additional notes during phenotyping without the need for further apps or handwritten remarks.

Partial selection of descriptors

It is possible to mark specific descriptors (square in the descriptor row, Figure 1) for selective data acquisition. Thus, the app only asks for indicated descriptors while all the others are skipped. This is particularly useful for phenotyping of multiple traits over time because it removes the necessity to manually skip descriptors not relevant at a specific time point.

Arrow keys displayed in the lower left corner

The arrow keys allow for a fast location jumping within a single row (up and down) or between rows (left and right).

Language

PhenoApp contains two language packages, English and German. The default language is English. If the global language setting of the mobile device is German, the language setting of the app changes automatically from English to German.

Important app settings

Zigzag mode

The app automatically jumps to the next listed descriptor after recording a data point, until the end of the descriptor list is reached. Then it moves directly to the first descriptor of the next location and so on. Logically, but quite unhandy in many practical cases, after reaching the end of a row, the app jumps to the first location in the next row (after the last row in a plot, it jumps to the next plot). To avoid walking unnecessary distances or the manual location correction, the app jumps in "zigzag" from the end of a row to the last location of the next row and counts backwards from there. This allows a continuous phenotyping while going through a field plot one row upwards and the next row backwards.

Arrow keys displayed in the top line give the direction of the next descriptor/location during phenotyping and are adaptable by a simple click on the button.

First empty button

This is a button on the top line that allows the user to jump to the first/next empty entry in their records. This is particularly useful for phenotyping multiple traits over time because it allows a fast tracking of missing data points.

Accession, passport, genotype, parents, collection no. and information

Displays the respective entry from the input file in the app if data are given.

Left-handed mode

The left-handed mode swaps the left and right side of the display for individual preferences.

Multiple selection

If the descriptor is of type 'rating', this option allows the user to select multiple values for a single trait. Results are separated with a comma in the output file. The activation of this setting removes the automatic jumping of the app to the next descriptor/location.
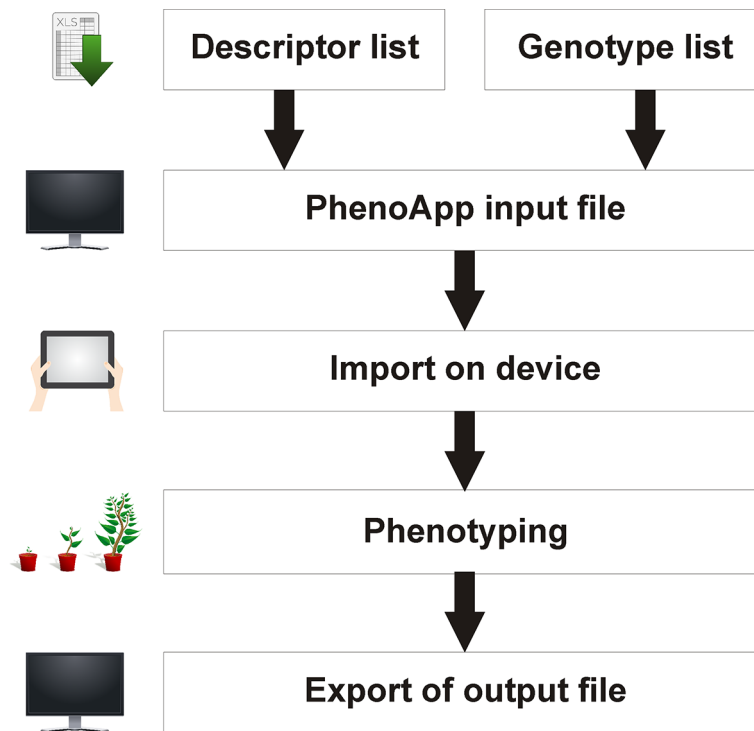
Figure 2. Principal PhenoApp workflow. A combined genotype-location and descriptor list is required as input. Post phenotyping, recorded data can be exported as a simple Microsoft Excel file. Minimal requirements are the PhenoApp on a mobile Android-based device and a computer with a spreadsheet software for data handling.

Operation

PhenoApp was developed with Java 8 using Google's appcompat-v7 and Apache's POI libraries. It works on mobile phones and tablets with android API-level 14, which stands for android version 4.0 (Ice Cream Sandwich), or higher and requires no internet connection. The software is open source under Apache 2.0 license (see *Software availability*[36]).

In general, PhenoApp is built as a simple interface to record phenotypic data. The app displays imported Microsoft Excel files containing a genotype and descriptor list and allows a data recording that can be saved and exported as Excel file for further use (Figure 2).

Use cases

Default

For most cases, PhenoApp can/will be used to digitally support phenotypic data acquisition in the course of specific projects or routine work. It only needs a computer with Microsoft Excel and a plant list (incl. locations), a defined scientific question that can be translated into a descriptor, and a mobile Android-based device. The data integration in superordinate structures or the descriptor adaptation according FAIR principles are desirable and emphasized, but not initially required.

Integration into local databases

For more than 10 years, IPK and the Julius Kühn Institute (JKI) have implemented a laboratory information management system (LIMS) as a universal platform for documenting and recording field, laboratory and bioinformatics data in order to establish a comprehensive system for research data documentation (reference Ghaffar *et al.* 2020; nr 32).

From this starting point, in a collaboration of IPK and JKI, PhenoApp was integrated in the LIMS workflow to advance data acquisition by allowing a fast, user friendly and direct import and export of datasets without the need of further modifications (Figure 3). Using the well-defined, file based data exchange interfaces of PhenoApp, we embed the creation of necessary import files into the LIMS. Even a tight integration for an optimal data re-usability and interoperability was straightforward to implement, for example the storage and versioning of scoring methods. This, in turn, enabled mapping of scoring schemes for the same species across observation cycles and projects by linking various
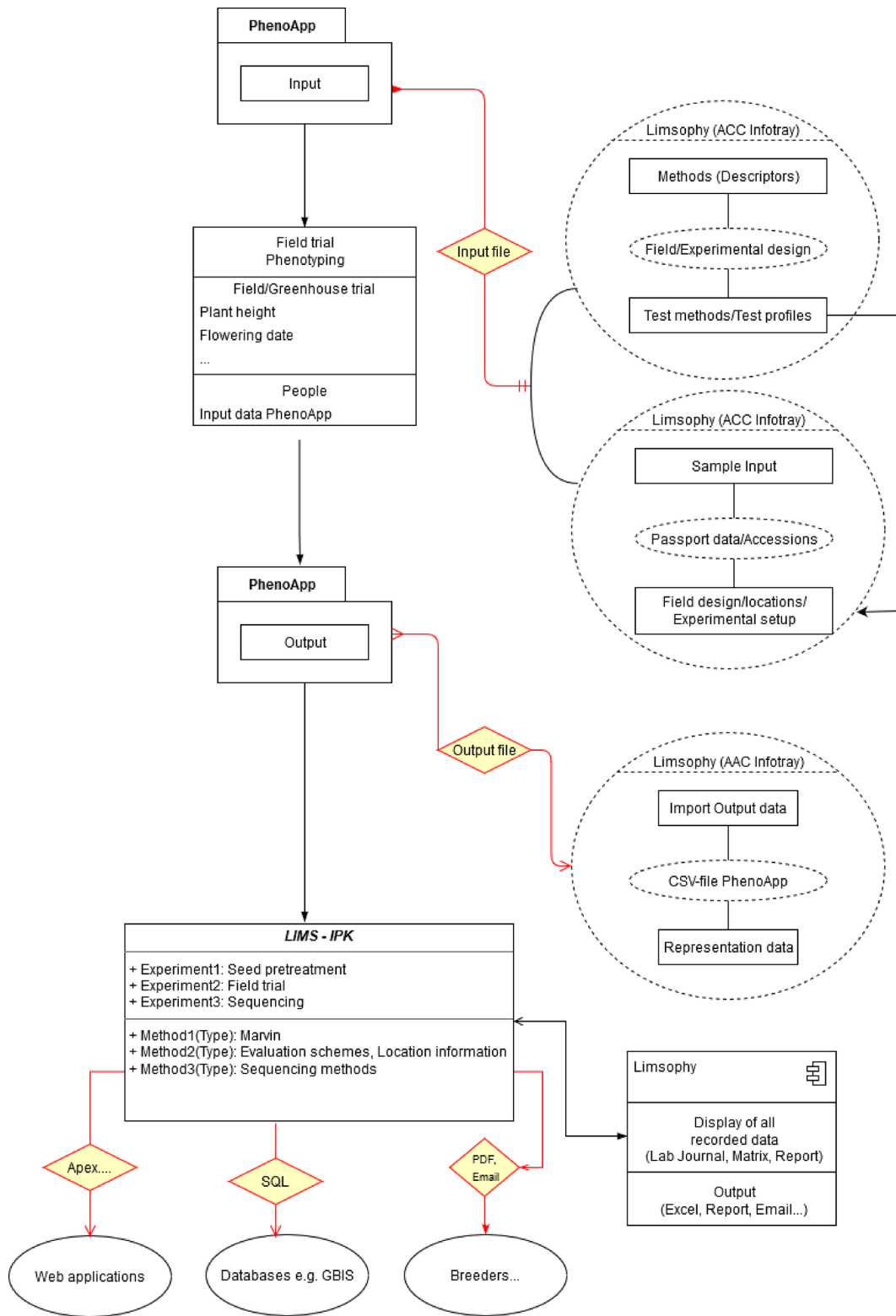
Figure 3. Data capture and integration into the laboratory information management system (LIMS) infrastructure of the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK).

Figure 4. Overview of the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) intern Apex (Oracle Application Express) application. Here, users will find a compilation of the data already registered in the laboratory information management system (LIMS), i.e. given descriptors and scales of the genebank as well as additional information such as specific contact persons. This offers a wide range of applications and enables quick and clear access to existing data and information for the development of new experiments and routine work.

rating schemes with all scoring schema details and even reference images. In addition, a user-friendly input mask in the system provides the possibility of precisely documenting the locations of the corresponding plants. This enables the re-use and thus optimal interoperability across field phenotyping studies.

The data recorded by PhenoApp are exported as a well-structured CSV file and pictures taken stored into a specified folder. When a phenotyping run is finished, this enables an ad-hoc and consistent import into the LIMS database. The potential of such FAIR enabled field phenotyping is the pre-cursor to feed Web information systems (Figure 4), data publication pipelines and APIs.[32] Currently, PhenoApp is used at the IPK within the barley pan-genome research project "SHAPE II" (https://shape.ipk-gatersleben.de/). Based on a defined so-called core set,[33] representative barley genotypes were selected, grown in different years at IPK and scored with PhenoApp.

Conclusions

PhenoApp provides a convenient and free-of-charge interface for a manual, digital and mobile data acquisition of a diverse range of phenotyping data. The user can easily create and customize their own descriptors and phenology scales. A simple entry to the application is possible for every user with basic knowledge of handling an Android device and spreadsheet software like Excel, without the need of integration in data management systems, although this is

possible as demonstrated. Adaptations according to FAIR data principles[17] or international information standards like MIAPPE[20] or EURISCO[34] are possible without any restrictions. Furthermore, the resulting input and output files can easily be integrated in workflows of data management systems. The benefits can range from a facilitated project-specific phenotyping in the short term to a support during routine work in the medium term towards a complete harmonization of an institutional wide data infrastructure in the long term. In particular, datasets that have been uniformly collected and stored for years in genebanks or breeding facilities bear a huge potential to accelerate the efficient use of genetic resources and therefore the development of new adapted crop varieties that ensure future food security.

Thinking further, PhenoApp was initially developed for, but is not limited to plant phenotyping. The app is applicable whenever a manual data recording is required and a specific location is given. An input format of three separated parts (plot – row – individual) is required, but can easily be adapted or filled with placeholders to expand the application range.

Data availability
Underlying data
Zenodo: PhenoApp Input/Output files. https://doi.org/10.5281/zenodo.5760899.[35]

This project contains the following underlying data:

- Input_example.xls (sample input file read in by PhenoApp. In addition, after installation, a sample input file will automatically be copied to the 'in' folder of the app main directory and no additional source data is required).

- Output_example.xls (sample output file created by PhenoApp).

Data are available under the terms of the Creative Commons Zero "No rights reserved" data waiver (CC0 1.0 Public domain dedication).

Software availability
Source code available from: https://gitea.julius-kuehn.de/JKI/pheno-app

Archived source code at time of publication: https://doi.org/10.5281/zenodo.5525779[36]

License: Apache-2.0

Acknowledgements

References

1. Yu X, Li X, Guo T, *et al.*: **Genomic prediction contributing to a promising global strategy to turbocharge gene banks.** *Nat Plants.* 2016; **2**: 16150.
PubMed Abstract | **Publisher Full Text**

2. Hickey LT, Hafeez AN, Robinson H, *et al.*: **Breeding crops to feed 10 billion.** *Nat. Biotechnol.* 2019; **37**(7): 744–754.
PubMed Abstract | **Publisher Full Text**

3. Nguyen GN, Norton SL: **Genebank Phenomics: A Strategic Approach to Enhance Value and Utilization of Crop Germplasm.** *Plants (Basel).* 2020; **9**(7).
Publisher Full Text

4. Norton SL, Khoury CK, Sosa CC, *et al.*: **Priorities for enhancing the ex situ conservation and use of Australian crop wild relatives.** *Aust. J. Bot.* 2017; **65**(8): 638–645.
Publisher Full Text

5. Pilling D, Bélanger J, Diulgheroff S, *et al.*: **Global status of genetic resources for food and agriculture: challenges and research needs.** *Genet. Res.* 2020; **1**(1): 4–16.

6. Tadesse W, Sanchez-Garcia M, Assefa SG, *et al.*: **Genetic Gains in Wheat Breeding and Its Role in Feeding the World.** *Crop Breeding, Genetics and Genomics.* 2019; **1**(1): e190005.

7. Furbank RT, Tester M: **Phenomics–technologies to relieve the phenotyping bottleneck.** *Trends Plant Sci.* 2011; **16**(12): 635–644.
PubMed Abstract | **Publisher Full Text**

8. Houle D, Govindaraju DR, Omholt S: **Phenomics: the next challenge.** *Nat. Rev. Genet.* 2010; **11**(12): 855–866.
PubMed Abstract | **Publisher Full Text**

9. Morrell PL, Buckler ES, Ross-Ibarra J: **Crop genomics: advances and applications.** *Nat. Rev. Genet.* 2012; **13**(2): 85–96.
Publisher Full Text

10. Rebetzke GJ, Jimenez-Berni J, Fischer RA, *et al.*: **Review: High-throughput phenotyping to enhance the use of crop genetic resources.** *Plant Sci.* 2019; **282**: 40–48.
PubMed Abstract | **Publisher Full Text**

11. Yang WN, Feng H, Zhang XH, *et al.*: **Crop Phenomics and High-Throughput Phenotyping: Past Decades, Current Challenges, and Future Perspectives.** *Mol. Plant* 2020; **13**(2): 187–214.
PubMed Abstract | **Publisher Full Text**

12. Kicherer A, Herzog K, Bendel N, *et al.*: **Phenoliner: A New Field Phenotyping Platform for Grapevine Research.** *Sensors-Basel.* 2017; **17**(7).
PubMed Abstract | **Publisher Full Text**

13. Tardieu F, Cabrera-Bosquet L, Pridmore T, *et al.*: **Plant Phenomics, From Sensors to Knowledge.** *Curr. Biol.* 2017; **27**(15): R770–R783.
PubMed Abstract | **Publisher Full Text**

14. Haucke T, Herzog K, Barre P, *et al.*: **Improved optical phenotyping of the grape berry surface using light-separation and automated RGB image analysis.** *Vitis* 2021; **60**(1): 1–10.

15. Rist F, Gabriel D, Mack J, *et al.*: **Combination of an Automated 3D Field Phenotyping Workflow and Predictive Modelling for High-Throughput and Non-Invasive Phenotyping of Grape Bunches.** *Remote Sens-Basel.* 2019; **11**(24).
Publisher Full Text

16. Zhao CJ, Zhang Y, Du JJ, *et al.*: **Crop Phenomics: Current Status and Perspectives. Frontiers.** *Plant Sci.* 2019; **10**.
Publisher Full Text

17. Wilkinson MD, Dumontier M, Aalbersberg IJ, *et al.*: **Comment: The FAIR Guiding Principles for scientific data management and stewardship.** *Sci. Data.* 2016; **3**: 160018.
PubMed Abstract | **Publisher Full Text**

18. Tian ZX, Wang JW, Li JY, *et al.*: **Designing future crops: challenges and strategies for sustainable agriculture.** *Plant J.* 2021; **105**(5): 1165–1178.
PubMed Abstract | **Publisher Full Text**

19. Watt M, Fiorani F, Usadel B, *et al.*: **Phenotyping: New Windows into the Plant for Breeders.** *Annu. Rev. Plant Biol.* 2020; **71**: 689–712.
PubMed Abstract | **Publisher Full Text**

20. Papoutsoglou EA, Faria D, Arend D, *et al.*: **Enabling reusability of plant phenomic datasets with MIAPPE 1.1.** *New Phytol.* 2020; **227**(1): 260–273.
PubMed Abstract | **Publisher Full Text**

21. Caracciolo C, Stellato A, Morshed A, *et al.*: **The AGROVOC Linked Dataset.** *Semant. Web.* 2013; **4**(3): 341–348.
Publisher Full Text

22. Smith B, Ashburner M, Rosse C, *et al.*: **The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration.** *Nat. Biotechnol.* 2007; **25**(11): 1251–1255.
PubMed Abstract | **Publisher Full Text**

23. Rocca-Serra P, Brandizi M, Maguire E, *et al.*: **ISA software suite: supporting standards-compliant experimental annotation and enabling curation at the community level.** *Bioinformatics.* 2010; **26**(18): 2354–2356.
PubMed Abstract | **Publisher Full Text**

24. Srivastava DJ, Vosegaard T, Massiot D, *et al.*: **Core Scientific Dataset Model: A lightweight and portable model and file format for multi-dimensional scientific data.** *PLoS One.* 2020; **15**(1):

e0225953.
PubMed Abstract | **Publisher Full Text**

25. Germeier CU, Unger S: **Modeling Crop Genetic Resources Phenotyping Information Systems.** *Front. Plant Sci.* 2019; **10**.
Publisher Full Text

26. Endresen DTF, Knüpffer H: **The Darwin Core extension for genebanks opens up new opportunities for sharing genebank datasets.** *Biodivers. Inform.* 2012; **8**.
Publisher Full Text

27. Meier U: *Growth stages of mono-and dicotyledonous plants.* Blackwell Wissenschafts-Verlag; 1997.

28. Oppermann M, Weise S, Dittmann C, *et al.*: **GBIS: the information system of the German Genebank.** *Database-Oxford.* 2015.

29. Aquino A, Barrio I, Diago MP, *et al.*: **vitisBerry: An Android-smartphone application to early evaluate the number of grapevine berries by means of image analysis.** *Comput. Electron. Agric.* 2018; **148**: 19–28.
Publisher Full Text

30. Komyshev E, Genaev M, Afonnikov D: **Evaluation of the SeedCounter, A Mobile Application for Grain Phenotyping.** *Front. Plant Sci.* 2017; **7**.
PubMed Abstract | **Publisher Full Text**

31. Müller-Linow M, Wilhelm J, Briese C, *et al.*: **Plant Screen Mobile: an open-source mobile device app for plant trait analysis.** *Plant Methods.* 2019; **15**: 2.
PubMed Abstract | **Publisher Full Text**

32. Ghaffar M, Schüler D, König P, *et al.*: **Programmatic Access to FAIRified Digital Plant Genetic Resources.** *J. Integr. Bioinform.* 2020; **16**(4).
PubMed Abstract | **Publisher Full Text**

33. Jayakodi M, Padmarasu S, Haberer G, *et al.*: **The barley pan-genome reveals the hidden legacy of mutation breeding.** *Nature.* 2020; **588**(7837): 284–289.
PubMed Abstract | **Publisher Full Text**

34. Weise S, Oppermann M, Maggioni L, *et al.*: **EURISCO: The European search catalogue for plant genetic resources.** *Nucleic Acids Res.* 2017; **45**(D1): D1003–D1008.
PubMed Abstract | **Publisher Full Text**

35. Schreiber T, Röckel F: **PhenoApp Input/Output files (1.0) [Data set].** *Zenodo.* 2021.
Publisher Full Text

36. Schreiber T, Röckel F: **PhenoApp (1.0).** *Zenodo.* 2021.
Publisher Full Text

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 2

Reviewer Report 04 January 2023

https://doi.org/10.5256/f1000research.141539.r156817

✔   **Andrew P. French**

University of Nottingham, Nottingham, UK

Thank you for the clear response to my comments - I am happy to recommend an approved rating for the article now.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Automated image analysis in plant phenotyping

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 13 December 2022

https://doi.org/10.5256/f1000research.141539.r156818

✔   **Karin Köhl** 🆔

Max Planck Institute of Molecular Plant Physiology, Potsdam, Germany

I thank the authors for their response and declare that I am satisfied with the information given. My only remaining request is that the recommendation to use a script to produce the input files is included in the conclusions.

Competing Interests: My co-worker and I have published a similar tool in 2015. I mentioned the reference to the authors to declare the potential conflict.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 29 July 2022

https://doi.org/10.5256/f1000research.77970.r140861

?

Andrew P. French
University of Nottingham, Nottingham, UK

Dear authors,

This paper describes a potentially useful tool to address the issue of manual data collection from plant phenotyping experiments. Such a tool, if well executed, goes some way to providing for an information management system for collecting and recording phenotyping data in the field. I have more experience with automated data collection though analysis of images, but I do recognise the need and utility of a tool to assist with data capture by hand. The app seems to focus almost entirely on data collection, and as noted in a response to a previous review, this data can then be fed to further downstream processing outside the app as required. The app therefore seems to be a semi-constrained front end for entering data into an Excel/CSV format. Whilst a simple idea, I believe this is useful in practice. I have some comments below.

It is a shame that there is no enforcing of controlled vocabulary on some of the fields. In the previous response to authors it is noted that the use of internationally-recognised descriptors is recommended, so not having the option to enforce this in some scenarios feels like a missing feature which should be present. Related, a lot of emphasis is put on the input file, which I believe you use a lab information system to supply. How well practically do you see this working if the institute does not have a LIMS?

I agree with the previous reviewer raising questions about Excel as a choice of data format. I think there are issues with this, and perhaps more flexible formats like JSON should be explored in the future (as you mention). However, there is no doubt Excel provides accessibility and readability of data in a form most phenotyping researchers will be familiar with.

The option of uploading images is welcome, especially as I can see the benefit in later using these alongside the manually collected data to support development of AI-based systems which can analyse the images. However, only allowing one image at a time seems like an oversight here. I am convinced it would be beneficial to allow users to upload as many images per entry as desired.

Is the rationale for developing the new software tool clearly explained?
Yes

Is the description of the software tool technically sound?
Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?
Partly

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?
Yes

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Automated image analysis in plant phenotyping

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

> Author Response 14 Sep 2022
> **Franco Röckel**, Institute for Grapevine Breeding Geilweilerhof, Siebeldingen, Germany
>
> First of all, we would like to thank the reviewer who gave us constructive comments and the opportunity to elaborate on certain aspects.To the review:
>
> **"Dear authors,**
> **This paper describes a potentially useful tool to address the issue of manual data collection from plant phenotyping experiments. Such a tool, if well executed, goes some way to providing for an information management system for collecting and recording phenotyping data in the field.**
> **I have more experience with automated data collection though analysis of images, but I do recognise the need and utility of a tool to assist with data capture by hand. The app seems to focus almost entirely on data collection, and as noted in a response to a previous review, this data can then be fed to further downstream processing outside the app as required. The app therefore seems to be a semi-constrained front end for entering data into an Excel/CSV format. Whilst a simple idea, I believe this is useful in practice. I have some comments below.**
> **It is a shame that there is no enforcing of controlled vocabulary on some of the fields. In the previous response to authors it is noted that the use of internationally-**

recognised descriptors is recommended, so not having the option to enforce this in some scenarios feels like a missing feature which should be present."

We agree that a strict control of vocabulary and normed descriptors would be of outstanding importance. Nevertheless, the tool reflects the unsatisfactory situation that stakeholder, like funding agencies, publisher or international core database, implemented no mechanisms to ensure this. Rather, it is moved to the level of informal commitments in institutional strategies. But even those need proper review. In order to provide a commonly usable data recording tool, the decision was to implement the data consistency in the database layer, which implements site-specific rules. On any inconsistency, a data steward is then responsible to resolve these. PhenoApp was built to be as much as possible customizable for maximum user friendliness. We do not want to enforce the use of any specific descriptors and consequently exclude users with individual (not yet harmonized) use cases. The controlled use of certain descriptors can easily be supervised on the institutional-wide level with read-only descriptor lists as explained in more detail in the next comment.
However, we recognize the possible benefit of such a feature. The integration of a common set of descriptors for specific species with a standardized vocabulary could be implemented in future updates. Steps to facilitate species-level descriptor management are the basis for such an option and are already under development.

"Related, a lot of emphasis is put on the input file, which I believe you use a lab information system to supply. How well practically do you see this working if the institute does not have a LIMS?"

At the beginning of the development, we did not use a LIMS either. We once created a common input file with a read-only descriptor list stored in an institutional-intern accessible folder. Therefore, users always used the same descriptor template and only added their specific plant list. It is more or less the same process as with a LIMS with a few more manual steps per user and without the automated LIMS database upload of the results after phenotyping. Since we use institutional-wide, standardized plant information (accession name, accession number, plant positions in the field etc.), the input files were completely standardized and the gathered information could be easily uploaded after the connection to our LIMS.
To conclude, we already established the standardized, institutional-wide use before the implementation of a LIMS. This should be feasible for every institute/lab in comparable manners as described above.

"I agree with the previous reviewer raising questions about Excel as a choice of data format. I think there are issues with this, and perhaps more flexible formats like JSON should be explored in the future (as you mention). However, there is no doubt Excel provides accessibility and readability of data in a form most phenotyping researchers will be familiar with."

Indeed the use of JSON as standardized, well-defined file format would have simplified the development of PhenoApp and ensured a robust file un-marshalling and data structure consistency per-se. As argued before, one aim is to support a wide range of application

areas and institutional infrastructures where some need to support Excel as widely used and accessible solution for data acquisition, editing and exchange. We saw the risk of JSON in losing those data maintainers and producers, who rely on proprietary historic database backend systems and data management processes. E.g. some do not use databases at all and archive in folder systems of Excel files as primary storage.

"The option of uploading images is welcome, especially as I can see the benefit in later using these alongside the manually collected data to support development of AI-based systems which can analyse the images. However, only allowing one image at a time seems like an oversight here. I am convinced it would be beneficial to allow users to upload as many images per entry as desired."

Please allow to clarfiy a potential misunderstanding. Only for a descriptor it is possible to upload one sample image. However, this should be sufficient to display representative images of rating schemes, phenological stages etc. in the app interface at the time of rating. Instead, for specific plant entries, it is not possible to upload images at all. However, we implemented the connection to the photo app of the devices used to take location-specific photos. There is no number limit for images per location here and naming is standardized based on the specified location, the actual timestamp and the variety name (if given) to ensure subsequent recognition. This allows clear allocation of the image to location/variety and if rating was performed for only one/few traits a clear allocation is possible even for the trait. Of course, these photos can be used for all downstream analysis without any restrictions.

Competing Interests: No competing interests were disclosed.

Reviewer Report 25 January 2022

https://doi.org/10.5256/f1000research.77970.r119127

? **Karin Köhl** (iD)

Max Planck Institute of Molecular Plant Physiology, Potsdam, Germany

Dear Authors,

Your manuscript describes a very useful tool for manual phenotyping based on mobile devices. However, the information on the tool need some amendments, especially for a user who is not familiar with the Limsophy-DB and wants to use it independently. I have some experience with using mobile devices for high-throughput phenotyping as we designed a similar tool for WinCE in 2015 (see reference).

Is the plant identification done by selection from a pre-existing list or by free text entry? In the latter case, how do you make sure the plant exists in the database/LIMS? How do you deal with duplication of the location names: two independent experiments may use the same combination of plot row and plant I.D? If the system is used independent of a LIMS system: How do you ensure that controlled vocabulary is used for species/variety? And, how do you ensure that controlled vocabulary for descriptors are used in the entry excel file.

Using excel files as an entry format has risks. How do you avoid Excel reformatting numbers into dates? How do you avoid users messing up the file by accidentally including carriage returns and other control characters?

For the numeric data entry: where do you define the unit in which the measurement was done (e.g. m, cm for height)?

How many clicks do you need to enter the BBCH (Biologische Bundesanstalt für Land- und Forstwirtschaft, Bundessortenamt und CHemische Industrie) information? Can you just enter the code or do you have to select from a list? The latter is very time consuming when you have to score a large number of plants.

How do you deal with multiple pictures of one BBCH stage, does the new image overwrite the old image or is it added?

Can you store the specific selection of descriptors as a protocol for future use, to make sure that each time the same set of traits is phenotyped. Is it possible to store the protocol independent of the Limsophy-DB?

Can you provide a reference for the Limsophy-DB?

Output excel file: how do you record the meaning of Sb and Bl if the system is not linked to a LIMS system? How do you maintain the link between input and output file?

References
1. Köhl K, Gremmels J: A software tool for the input and management of phenotypic data using personal digital assistants and other mobile devices.*Plant Methods*. 2015; **11**: 25 PubMed Abstract | Publisher Full Text

Is the rationale for developing the new software tool clearly explained?
Yes

Is the description of the software tool technically sound?
Partly

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?
Partly

Is sufficient information provided to allow interpretation of the expected output datasets

and any results generated using the tool?
Partly

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?
Partly

Competing Interests: My co-worker and I have published a similar tool in 2015. I mentioned the reference to the authors to declare the potential conflict.

Reviewer Expertise: plant phenotyping, data management, abiotic stress tolerance

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 23 Feb 2022
Franco Röckel, Institute for Grapevine Breeding Geilweilerhof, Siebeldingen, Germany

First of all, we would like to thank the Reviewer who has provided us with many insightful comments and a chance to improve our work. To the review:

Your manuscript describes a very useful tool for manual phenotyping based on mobile devices. However, the information on the tool need some amendments, especially for a user who is not familiar with the Limsophy-DB and wants to use it independently. I have some experience with using mobile devices for high-throughput phenotyping as we designed a similar tool for WinCE in 2015 (see reference).

Is the plant identification done by selection from a pre-existing list or by free text entry?
The PhenoApp is designed for phenotyping of previously defined material with given locations (e.g. breeding material in field) that is screened with defined descriptors. This information is provided to the App in an Excel file as described under "Data import and export" prior to the phenotyping. In the "Locations" sheet, the material can be filled with the individual material, while in the sheet "Traits" the descriptors can freely be customised. Data to the Excel Import file can be (1) entered directly, (2) by copy & paste from pre-existing lists/files, or (3) from databases (like LIMS) using matching data export settings.

In the latter case, how do you make sure the plant exists in the database/LIMS?
There is no control via PhenoApp. If you use the PhenoApp independently of a database you can use the Excel output file with the data to feed every planned downstream application. For data import to a database/LIMS we highly recommend to use a script to create the input files and a second one that checks every entry during data import after phenotyping. This avoids errors *a priori* and gives full control over the clear allocation and completeness.

How do you deal with duplication of the location names: two independent

experiments may use the same combination of plot row and plant I.D?
Each location (or plant) is linked to an experiment, so there are no problems if a location appears in several experiments. Thus, only the associated locations are used in each experiment. If a specific location is given multiple times in a single list with different names, all entries are displayed in PhenoApp, although only one is true and others are wrong. Users will recognize it and can then adapt the input file.

If the system is used independent of a LIMS system: How do you ensure that controlled vocabulary is used for species/variety? And, how do you ensure that controlled vocabulary for descriptors are used in the entry excel file.
We designed PhenoApp to be as customisable as possible. There is no controlling of common vocabulary for any descriptor. Nevertheless, we recommend using international standardised descriptor lists as mentioned in the text. The validation of the correct taxonomy rely on a central defined vocabulary or ontology. This central place should be a database that offer access and validation features. Nevertheless, there are spreadsheet based terminology validation features, e.g. RightField (DOI: 10.1093/bioinformatics/btr312). Furthermore, one may even implement MS-Excel macros that validate to a remote accessible list of valid species.

Using excel files as an entry format has risks. How do you avoid Excel reformatting numbers into dates?
Indeed this is a potential pitfall as reported already in Browman et al. (10.1080/00031305.2017.1375989) By following these rules, there have been no problems of this kind with Excel so far. After phenotyping, dates were displayed correctly. During development, Excel seemed to be the simplest and most used format for scientists to define their experiments independently. In the future, it is planned to support JSON as an alternative input format. Furthermore, csv file format can be used for export instead of xlsx.

How do you avoid users messing up the file by accidentally including carriage returns and other control characters?
So far, no problems of this kind were known and no controlling system is used. PhenoApp will display the special characters. In our institutes, the input files are mainly created directly by the LIMS to avoid these errors and where this is not possible, standardised input file templates are available. To a certain degree, users are responsible for the correctness of their own datasets.

For the numeric data entry: where do you define the unit in which the measurement was done (e.g. m, cm for height)?
This can be specified in the descriptor itself. We recommend defining the measurement including the used unit as precisely as possible in the "Remark" field of the "Traits" sheet in the input file.

How many clicks do you need to enter the BBCH (Biologische Bundesanstalt für Land- und Forstwirtschaft, Bundessortenamt und CHemische Industrie) information? Can you just enter the code or do you have to select from a list? The latter is very time consuming when you have to score a large number of plants.
Every pre-integrated BBCH scale has three list levels: species, macro stage and micro stage.

Species and the macro stage need to be selected once in the beginning meaning that with every new plant entry both entries are kept. Only the micro stage needs obviously to be selected at every location entry. You need one click to open the micro stage menu and one additional click for the individual BBCH stage at every location. It is quite handy and approved.

How do you deal with multiple pictures of one BBCH stage, does the new image overwrite the old image or is it added?
There is only one image per stage/descriptor possible and a new image would replace the old one. If really two images are needed, users would have to merge them into one image file and then upload it. This has been done in the past.

Can you store the specific selection of descriptors as a protocol for future use, to make sure that each time the same set of traits is phenotyped. Is it possible to store the protocol independent of the Limsophy-DB?
We store our descriptor list in the LIMS database. This list is always used when an input file is created. Before we integrated the LIMS in the PhenoApp workflow, we created an empty input file (empty of plant locations) with a read-only descriptor list in Excel and saved it in an institutional-wide PhenoApp filesystem folder share for further use. Every new user was informed to rely on this standard list. Next to that, a user is free to adapt his individual descriptor list in the input file to only the relevant descriptors he is going to use in the given experiment or development state by deleting all others.

Can you provide a reference for the Limsophy-DB?
Ghaffar, M., Schuler, D., Konig, P., Arend, D., Junker, A., Scholz, U., Lange, M., 2020. Programmatic Access to FAIRified Digital Plant Genetic Resources. J Integr Bioinform; DOI: 10.1515/*jib-2019-0060*

Output excel file: how do you record the meaning of Sb and Bl if the system is not linked to a LIMS system?
A detailed meaning of every descriptor can be given in the remark column, which is a free text field. We recommend defining records of individual rating systems here as precisely as possible including units for measured values, phenological state, organ positions etc.

How do you maintain the link between input and output file?
The complete location list of the input file is also copied to your output file independent of your phenotyping.

Competing Interests: The reviewer mentioned an own tool published in 2015. We carefully checked the reference and agree that there might be a potential conflict to a certain degree.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias

- You can publish traditional articles, null/negative results, case reports, data notes and more

- The peer review process is transparent and collaborative

- Your article is indexed in PubMed after passing peer review

- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research