



Datenmanagementpläne zwischen Vorgaben der Förderer und Forschungspraxis

Tsu-Wei Chen

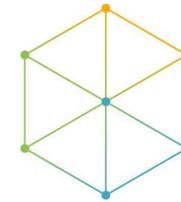
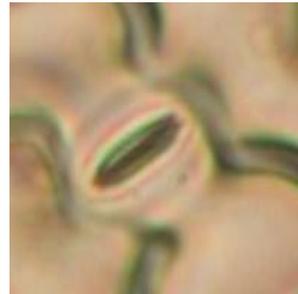
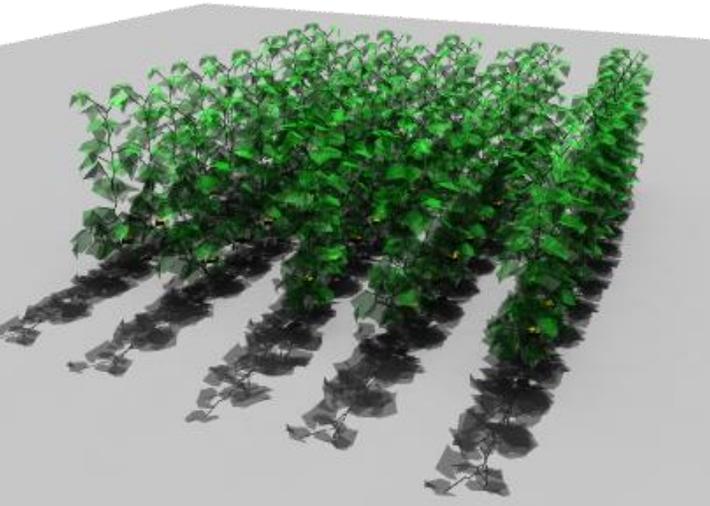
Professorship for Intensive Plant Food Systems

Lentzeallee 75, 14195, Berlin

Email: tsu-wei.chen@hu-berlin.de

ORCID: [0000-0001-6698-563X](https://orcid.org/0000-0001-6698-563X)

Twitter:  @TsuWeiChen1



CUBES Circle
Future Food Production



Bundesministerium
für Bildung
und Forschung

About my group and my project

Professor of Intensive Plant Food Systems + DFG Emmy Noether Program

HUMBOLDT-UNIVERSITÄT ZU BERLIN



Group leader: Prof. Dr. Tsu-Wei Chen

Senior scientists: PD Dr. Heiner Grüneberg, Dr. Ina Pinker & Dr. Susann Müller

PostDoc: Dr. Andries Temme (BMBF CUBESCircle)

PhD students: Tien-Cheng Wang (DFG) & Annette Mannschke (DFG Emmy Noether)

Senior technical assistant: Stefanie Sentner

Technical assistant: Constantin Adanczak (DFG Emmy Noether)

Student assistants: Clara Schuster, Sven Schweiter



Difficulties:

- **Generation gap (!!)**
- **New team**
- **Different projects**
- **Common language**

About my group and my project

Type of data and special challenges

Data type

Data levels

Chemical/
Genomics

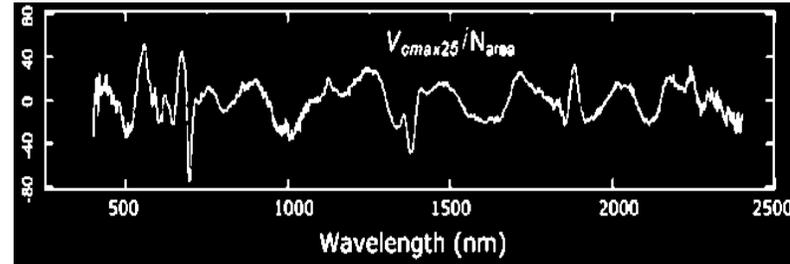
Organ

Plant

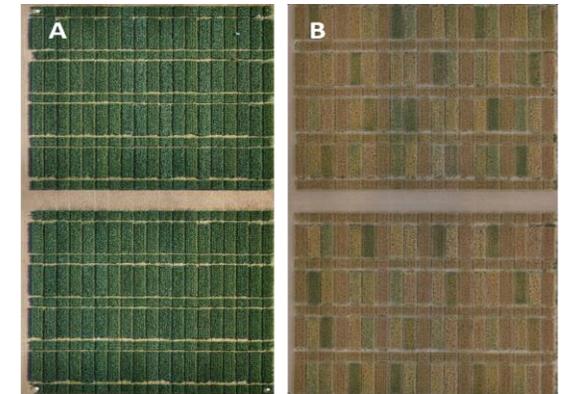
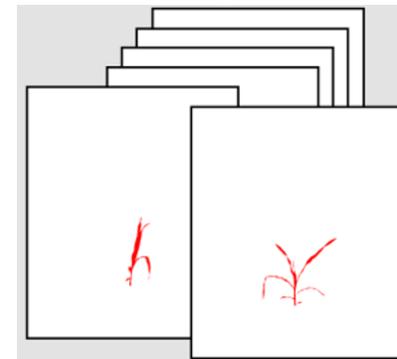
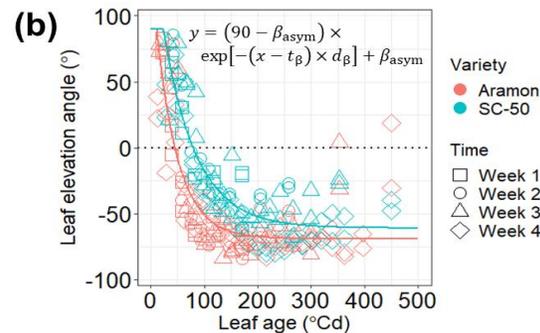
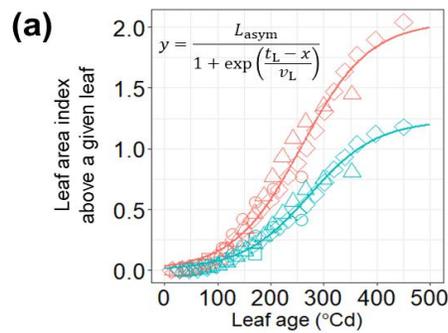
Canopy

Raw

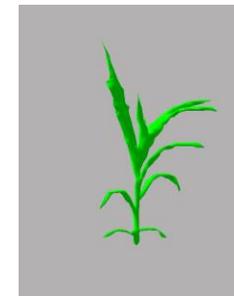
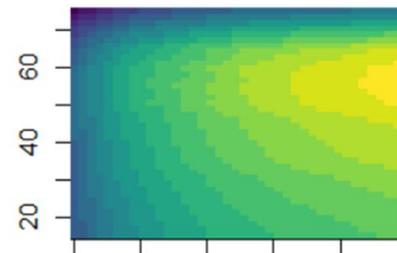
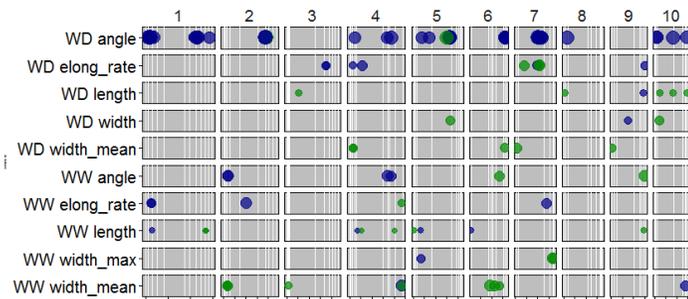
AGTCTGA
ATCCGGA
ACTGATC
TAGCTTT



Processed



Simulated





Umgang mit den im Projekt erzielten Forschungsdaten:

„Die Verbesserung des Umgangs mit Forschungsdaten hat sowohl bei nationalen und internationalen Forschungsorganisationen, als auch in der Wissenschaft eine hohe Bedeutung. Die DFG ist daher bestrebt, durch ihre Förderung auch zur

- 1) *Sicherung,*
- 2) *Aufbewahrung und*
- 3) *Nachnutzbarkeit beizutragen.*

Local/Cloud/Server/Repository

Data model/system

Wenn im Projekt systematisch Forschungsdaten oder Informationen gewonnen werden, *erläutern Sie bitte Art, Umfang und Dokumentation der Daten sowie die geplante Aufbewahrung*. Gehen Sie auch auf die Möglichkeit der Nachnutzung durch andere Wissenschaftlerinnen und Wissenschaftler ein. Bitte berücksichtigen Sie dabei auch - sofern vorhanden - die in Ihrer *Fachdisziplin existierenden Standards* und die *Angebote existierender Datenrepositorien oder Archive*. Weitere Anregungen und Best-Practice-Beispiele finden Sie unter <http://www.dfg.de/antragstellung/forschungsdaten/>

Die für die Nachnutzung der Forschungsdaten *anfallenden projektspezifischen Kosten können Sie im Rahmen des Projekts beantragen*. Stellen Sie in diesem Fall bitte auch dar, in welcher Form eine Unterstützung beim Daten- und Informationsmanagement durch die am Projekt beteiligten Institutionen geleistet wird. Steht die Nachnutzbarkeit der entstehenden Forschungsdaten in engem Zusammenhang mit Forschungsobjekten (Gewebe, Zelllinien, Installation, *Materialien*, Kunstgegenstände o.ä.) so bitten wir Sie, auch Angaben zur Aufbewahrung dieser Objekte im Antrag zu ergänzen“



Dezernat 4 Forschung und EU-Hochschulbüro,
Technologietransfer

Forschungsdatenmanagement
Dr. Volker Soßna

Allgemeine Beratung
Dr. Maria Leuteritz
Dr. Sonja Detay

➔ - Clear enough
- No question from the reviewers

2.4 Data handling

Data handling adheres to the [DFG Guidelines on the Handling of Research Data](#) and the [Guidelines for handling research data at Leibniz University Hannover](#). On approval, we will draft a data management plan (DMP) detailing the measures described in this chapter. The DMP will particularly define standards for file naming, the documentation of data processing and the recording of metadata. We will take into account existing standards and best practice examples, e.g. those elaborated by the [BonaRes](#) project or possibly in a discipline-specific NFDI consortium.

We expect the following types of data:

- photographs of plants for automated leaf area calculation etc. [Format?]
- tabular data from lab measurements (csv)
- R-scripts for data evaluation

Data not already stored in file formats suitable for long-term preservation will be converted in such formats before archiving or publishing.

During the project, participating researchers store a copy of the data they collect and process in a secure cloud storage provided by Leibniz University IT Services (LUIS). We will save additional copies to the server of the Institute of Horticultural Production Systems at LUH, which is backed up by LUIS on a daily bases. The cloud storage is part of the [service bundle "Projektablage"](#), which also contains a service for [automated synchronisation \(seafile\)](#) of local drives with the cloud.

We process the raw data with [InfraPhenoGrid](#), a scientific workflow infrastructure for plant phenomics on the grid. Statistical analyses will be carried out with R. For each file containing scientific data we systematically collect metadata in a [\[data base / master table / electronic lab book...\]](#). The metadata will be compatible with the [DataCite](#) Standard. We will also systematically collect more general information, e.g. on our research and experimental designs, in a [Wiki](#).

At the end of the project, we will publish relevant research data [open access](#) and under an open license ([CC0](#) or [CC-BY](#)) in either the data repository of LUH or a suitable discipline-specific repository if available, such as [one of the data centres of the GFBio consortium](#). Data may be publically accessible only after an embargo period of up 12 month if they are the bases for still unpublished journal articles. On upload, we will get a DOI for each dataset, which we will indicate in related scientific articles. We will store an additional copy of all project data, including those not chosen for publication, in the [LUIS data archive](#). The archive guarantees [bitstream preservation](#) for at least ten years.

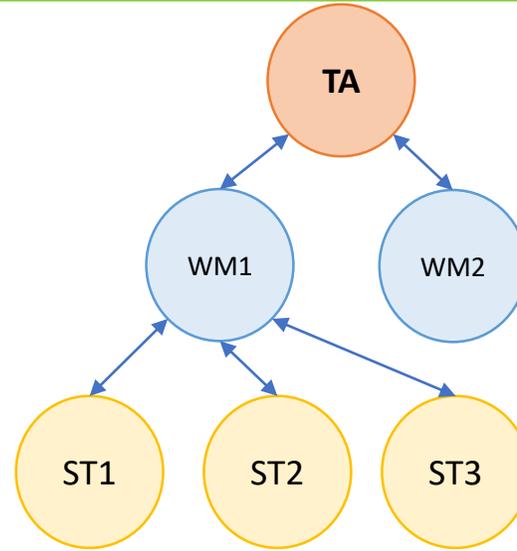
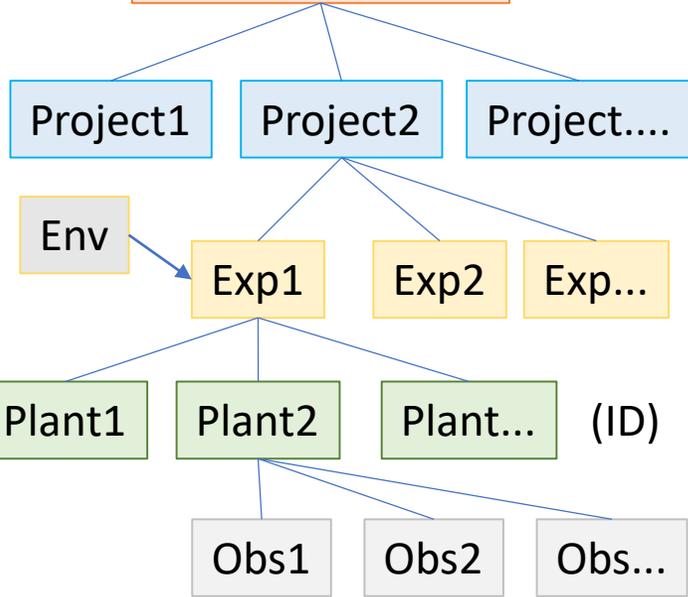
Whenever the need arises, the research data support team at Leibniz University Hannover will provide counsel and support on all aspects of data handling.

- Soßna, Volker Dieser Text versteht sich nur all
- Soßna, Volker Bitte beachten Sie auch die
- Soßna, Volker Wir haben dazu eine kleine
- Soßna, Volker Meine Empfehlung: orientieren
- Soßna, Volker Behalten Sie auch im Auge, wa
- Soßna, Volker Näheres dazu finden Sie hier:
- Soßna, Volker Der folgende Absatz müsste
- Soßna, Volker Das wäre sinnvoll, wenn Sie vi
- Soßna, Volker Ich meine mich zu erinnern, da
- Soßna, Volker Haben Sie sich überlegt, wo un
- Soßna, Volker Wenn das für Sie sinnvoll ist.
- Soßna, Volker Sie sollten Ihre zur
- Soßna, Volker Also ohne irgendwelche
- Soßna, Volker Es gibt sehr viele verschiedene
- Soßna, Volker Keine Rechte vorbehalten. Ist
- Soßna, Volker Keine Rechte vorbehalten auß
- Soßna, Volker <https://www.gfbio.org/data>.
- Soßna, Volker Damit wären Sie dann wirklich
- Soßna, Volker Das Archiv legt immer mehrer

In reality

How to achieve a good practice

Data hierarchy of a working group



Data curator

Scientific staff
Supervisor

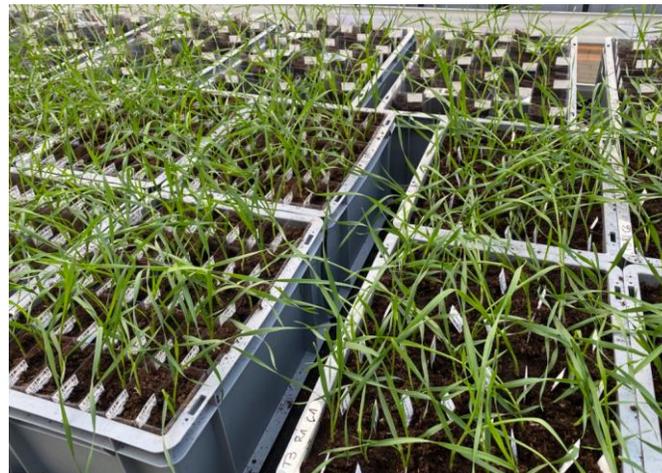
Technical assistant
Student assistant
BSc & MSc. student



- Cloud service (1TB @HU)
- Consistent folder structure in each experiment
- Quick check & double-check



- Creating meta-data (design & ID)
- Checking data quality by visualization
- Ensuring machine readability and reproducibility (analysis)



Difficulties:

- **Generation gap** → further training
- **New team** → training course
- **Different projects** → consistent data structure
- **Common language** → consistent understanding

Data management costs lots of time → dilemma of PhD student and early career scientists → curator needed

„PhD students and early career scientists need publications and data curation is rather a service (not credited) and is often undervalued“

Rolf Apweiler (EMBL-EBI Heidelberg)
11.03.2021 in Statusseminar Plant 2030

Question:

- 1) How can I afford a data curator for a small 3-year project?
- 2) Is a core-project in DFG research unit or BMBF consortium sufficient for data curating?
- 3) When should the data be open?

Success through a data curator



Bundesministerium
für Bildung
und Forschung



BRIWECS
wheat for tomorrow

ARTICLES

<https://doi.org/10.1038/s41477-019-0445-5>

nature
plants

Breeding improves wheat productivity under contrasting agrochemical input levels

Kai P. Voss-Fels^{1,2,11}, Andreas Stahl^{1,11}, Benjamin Wittkop^{1,11}, Carolin Lichthardt³, Sabrina Nagler⁴, Till Rose⁴, Tsu-Wei Chen³, Holger Zetzsche⁵, Sylvia Seddig⁶, Mirza Majid Baig⁷, Agim Ballvora⁷, Matthias Frisch⁸, Elizabeth Ross², Ben J. Hayes², Matthew J. Hayden⁹, Frank Ordon⁵, Jens Leon^{7,10}, Henning Kage⁴, Wolfgang Friedt^{1*}, Hartmut Stützel^{3*} and Rod J. Snowdon^{1*}

Thanks for your attention
and team/ cooperation partners / funding agencies

HUMBOLDT-UNIVERSITÄT ZU BERLIN



DFG
Deutsche
Forschungsgemeinschaft



INRAE

Inria
inventors for the digital world



JUSTUS-LIEBIG-
 UNIVERSITÄT
GIESSEN

Humboldt Universität zu Berlin

Dr. Andries Temme
Dr. Ina Pinker
Dr. Susann Müller
PD Dr. Heiner Grüneberg
Tien-Cheng Wang
Annette Mantschke
Constantin Adamczak
Stephanie Sentner

Leibniz Universität Hannover

Prof. Dr. Hartmut Stützel
Dr. Dany Moualeu-Ngangue
Yi-Chen Pao

INRAE/INRIA

Dr. Francois Tardieu
Dr. Christian Fournier
Dr. Christophe Pradal
Dr. Llorenç Cabrera-Bosquet

 @TsuWeiChen1

- Haben die FDM/DMP-Vorgaben/-Vorlagen der Förderer einen praktischen Nutzen für die Forschenden?
- Ja, eigentlich. Aber in Form einer Checkliste/Tabelle wäre noch besser
- Wie transparent ist die Begutachtung/Evaluation von DMP/FDM-Angaben? Wird die Umsetzung geprüft?
- Begutachtung (?). Wie wird die Umsetzung geprüft?
- Wie gut funktioniert die Kommunikation zwischen Forschung, Förderern und Beratungspersonal?
- Fragmentarisch