UPOV
International Union for the Protection of New Varieties of Plants

E

**Technical Working Party for Agricultural Crops**

**TWA/51/7**

**Fifty-First Session**
**Cambridge, United Kingdom, May 23 to 27, 2022**

**Original:** English
**Date:** May 5, 2022

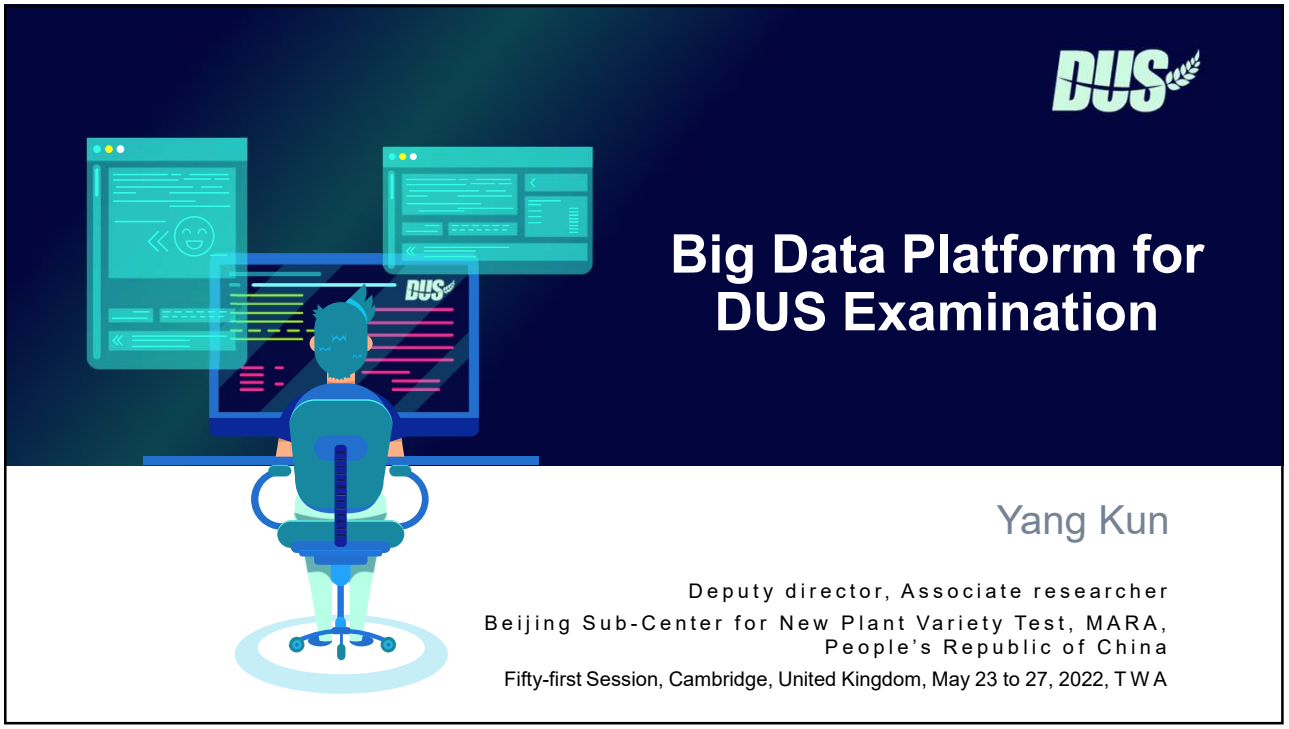
# BIG DATA PLATFORM FOR DUS EXAMINATION

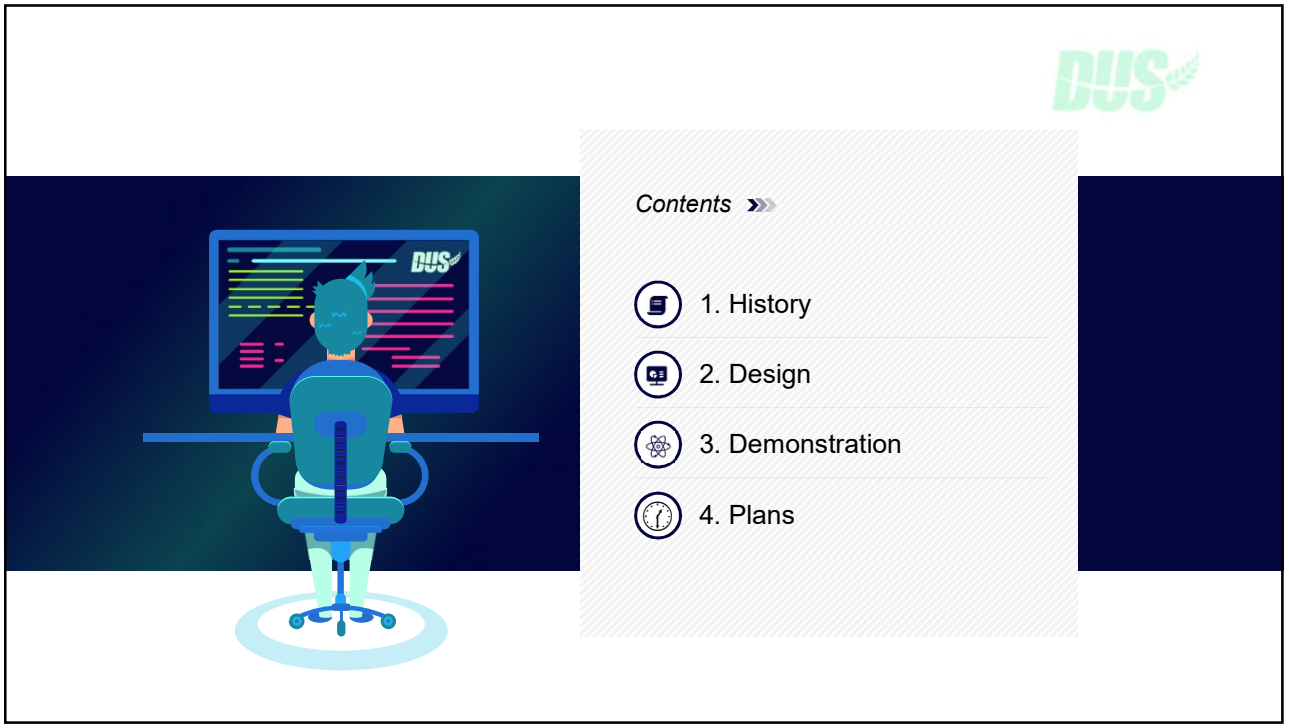
*Document prepared by an expert from China*

*Disclaimer: this document does not represent UPOV policies or guidance*

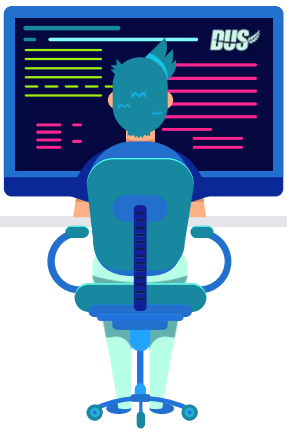
The annex to this document contains a copy of a presentation “Big Data Platform for DUS Examination”, to be made by an expert from China, at the fifty-first session of the TWA.

[Annex follows]


DUS
Big Data Platform for DUS Examination
Yang Kun
Deputy director, Associate researcher
Beijing Sub-Center for New Plant Variety Test, MARA, People's Republic of China
Fifty-first Session, Cambridge, United Kingdom, May 23 to 27, 2022, TWA


DUS
Contents
1. History
2. Design
3. Demonstration
4. Plans

# 01 History




## History

**2017-2018 v1.0**
electronic data recorder

**2019 v2.0**
management platform for data

**2020 v3.0**
deep design for DUS from task to report

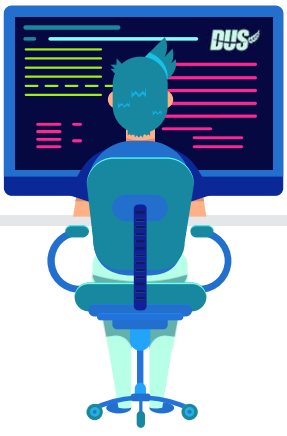
**2021 v4.0**
SAAS design, image analysis, statistical analysis

**2022 v5.0**
integrated system for DUS, VCU, DNA, chemical contents, resistance

## History

1. 2021, report on version 4.0 in TWC39.
2. 2022, report on version 5.0 in TWA51.

# 02 Design



## Problems of current DUS testing

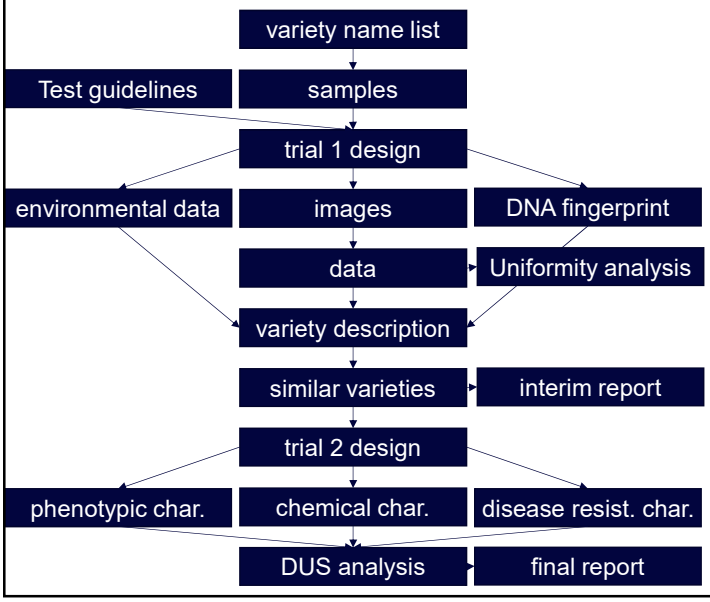
1. Field trial have long period, heavy workload, big influence by environment, big observation error between testers, low efficient analysis.
2. Molecular distance doesn’ t have good relationship with morphological distance.
3. Lack of harmonized theory or rule of calculation of minimum distance of D, U, and S which means same data always have different results in different stations or countries.

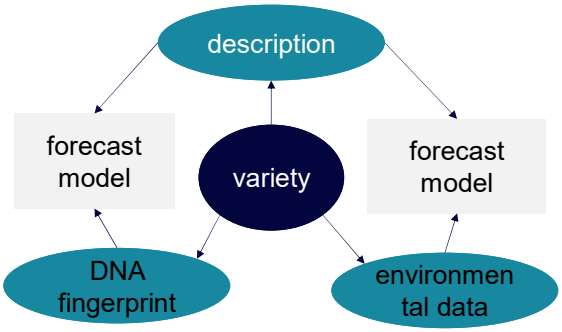
## How to store the big data?

- **Beijing station per year:**
  - **17 crops**
  - **more than 900 candidate varieties**
  - **more than 1,500 varieties**
  - **more than 3,060,000 data**
  - **more than 7,500 photos**
  - **more than 40,000 molecular data**
  - **7430 samples and descriptions, 37,150 photos, 4210 molecular data**
- **All stations per year:**
  - **more than 1,000 crops**
  - **more than 10,000 candidate varieties**
  - **more than 20,000 varieties**
  - **more than 40,800,000 data**
  - **more than 100,000 photos**
  - **more than 400,000 molecular data**
  - **more than 30,000 samples and descriptions, 150,000 photos, 10,000 molecular data**

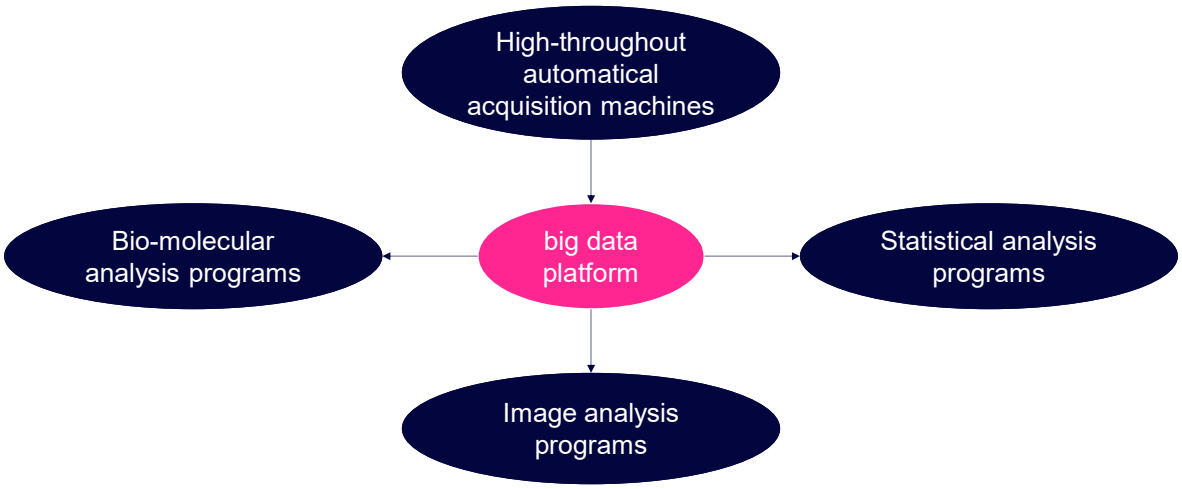
**Solution: big data platform based on SaaS structure**

## How to make result effective?

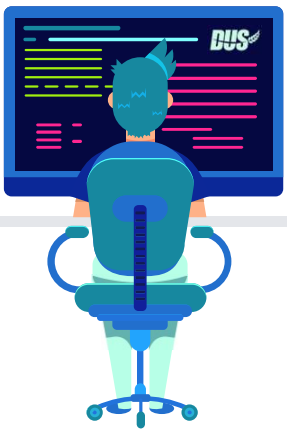


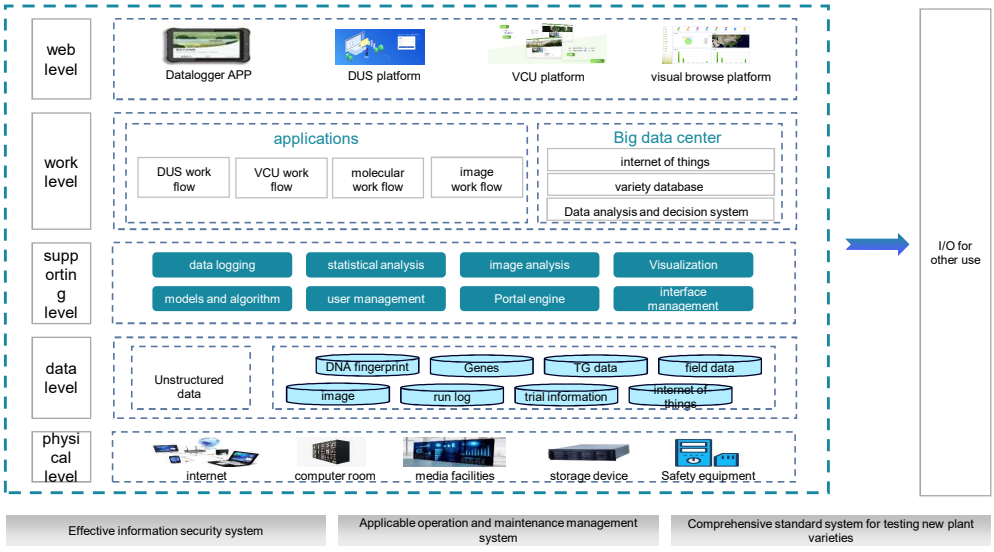

## How to make analysis efficient?



# 03 Demonstration



## Structure of whole system



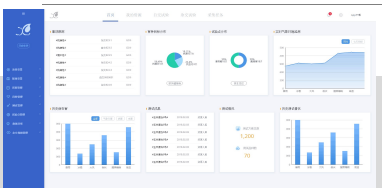
## DUS platform cloud+APP+AI+big data



stations and testers | TGs | varieties | tasks | samples | trials | data collection and analysis | reports

tester's platform

## VCU platform cloud+APP+AI+big data



### VCU platform

APP: Field datalogger


application of trial

examination plan

report

management of examination plan

1.import trial plans
2.review companies' plans

management of site plan

site plan

site data

1.choose site
2.designate plan
3.visit site
4.supervise data real-time

site work

1.collect data
2.submit data

1.submit trial plans
2.receive samples
3.supervise data real-time

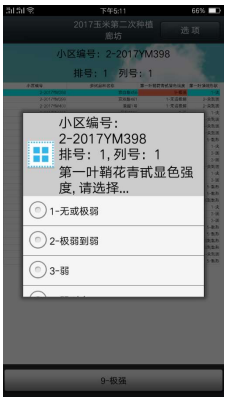
data analysis

1.summarize data(plan, variety, site)
2.analyze data (site quality analysis, ANOVA of trials, analysis of productivity, stability and suitability)

## Mobile data logger


select trial

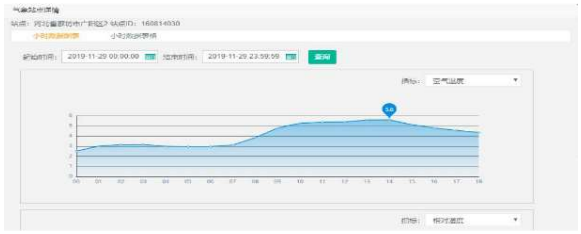

set record model




data record

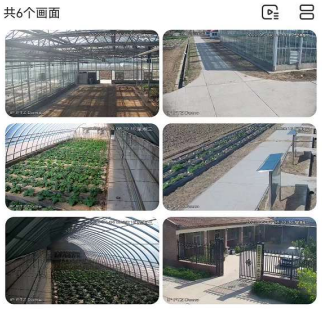
## Collection of environmental data

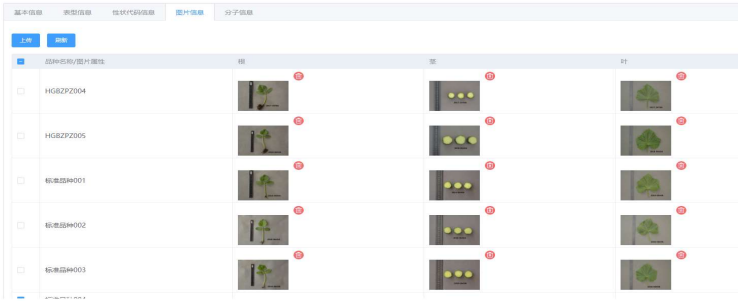
national data



station’s data

## Collection of image data by unmanned machines



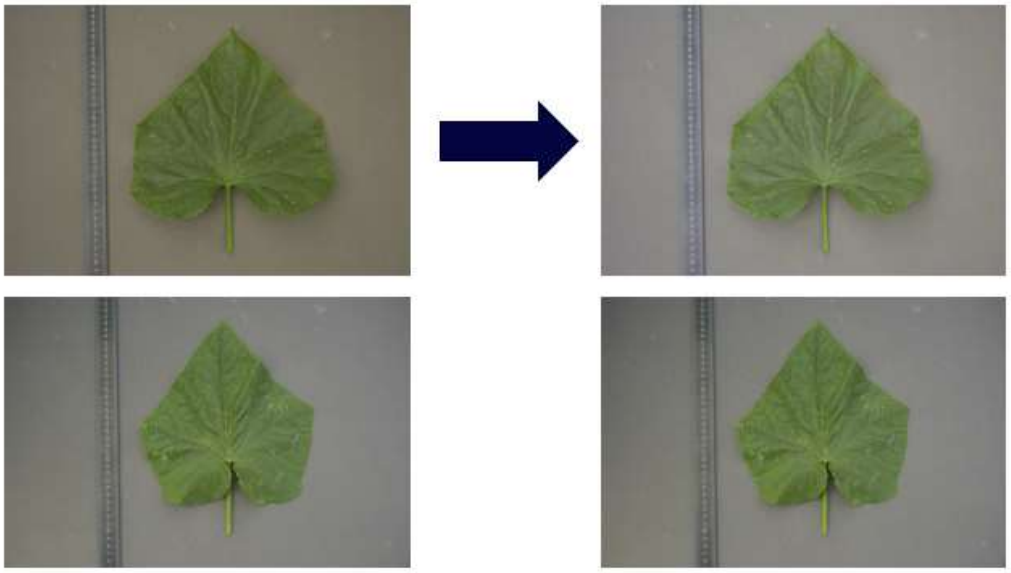
## Image analysis platform

分析部位: 幼苗 叶 果实

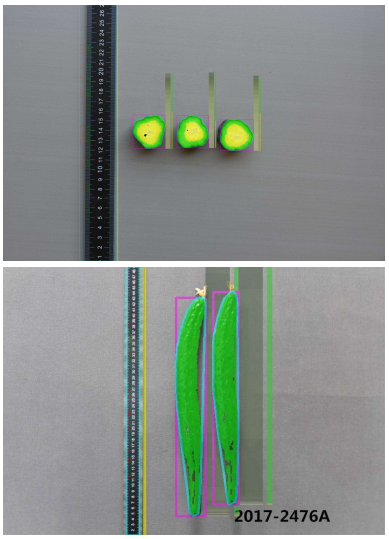
| 品种 | 品种类型 | 图片 | 分析结果图片 | 分析结果 |
|---|---|---|---|---|
| HG申请品种1 | 申请品种 |  |  | 图像名:aee8a40d-1d2f-49c6-a5d6-f322736f37ab.jpg;red:0;orange:171;yellow:17876;green:3870274;cyan:0;blue:0;purple:0;叶片长度（像素）:2938;叶片长度（mm）:303.277;叶片面积（平方mm）:41597.9 |
| HG申请品种2 | 申请品种 |  |  | 图像名:ddab3cf8-ed70-48ab-bd5a-311e3ce0e8b8.jpg;red:0;orange:171;yellow:17876;green:3870274;cyan:0;blue:0;purple:0;叶片长度（像素）:2938;叶片长度（mm）:303.277;叶片面积（平方mm）:41597.9 |

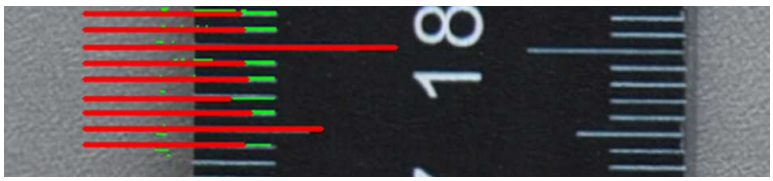
返回

## Color calibration

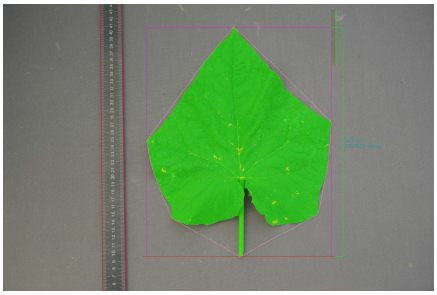


## Ruler calibration

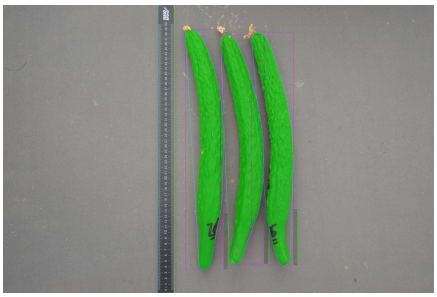


## Image analysis of cucumber leaf

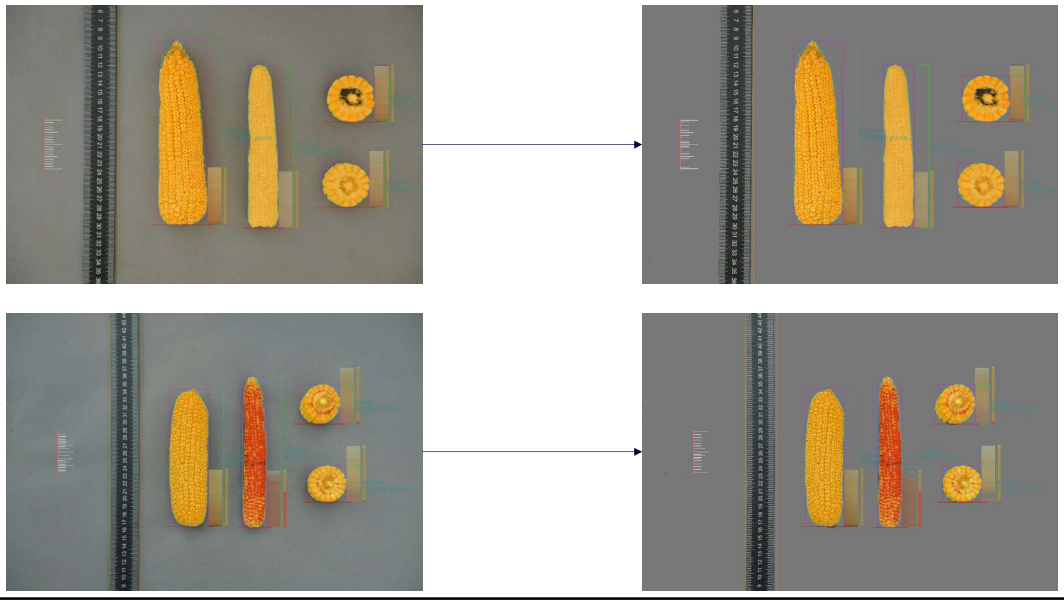


- Automatically recognize the size of the ruler to calculate the size of the leaf
- leaf area, width, length, convex hall features, angle of the tip, intensity of dentation of margin, shape of leaf, shape of apex of terminal lobes intensity of green color etc..
- Exploring using AI to automatically dig more traits difficult for human to describe.

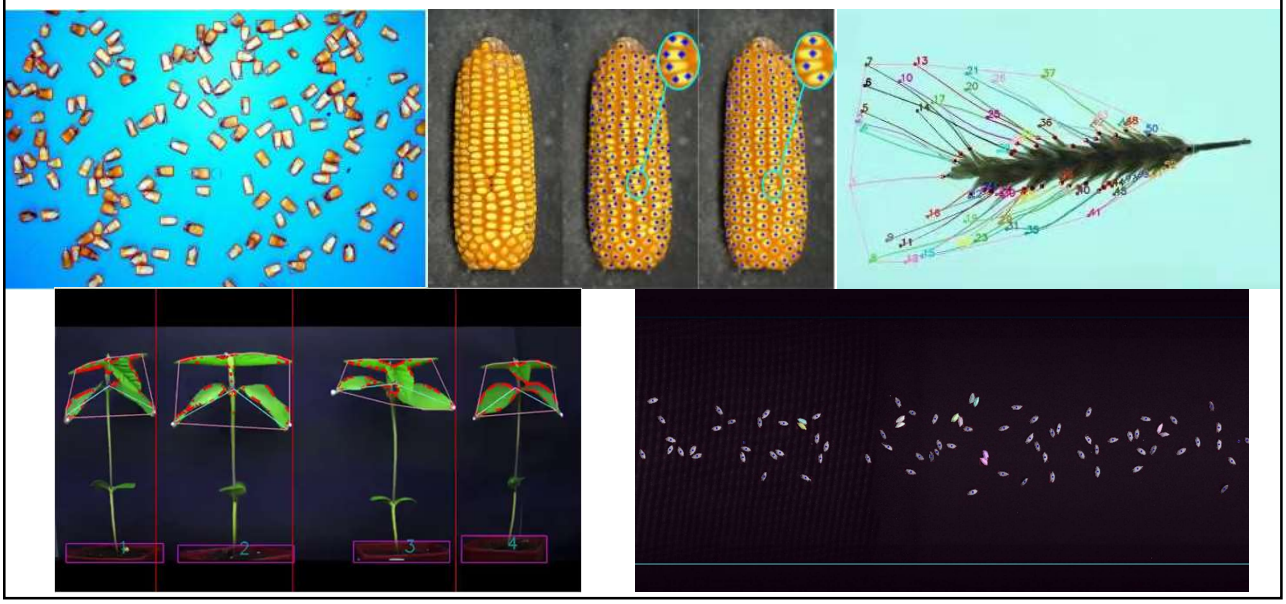


- Automatically recognize the size of the ruler to calculate the size of the fruit.
- fruit area, width, length, shape of calyx end, length of stem end, ground color of skin, secondary color of skin, distribution of secondary color, intensity of glossiness, density of vestiture etc..
- Exploring using AI to automatically dig more traits difficult for human to describe.

## Image analysis of Maize

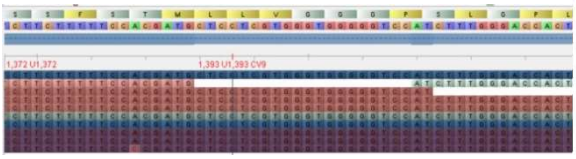


## Other image analysis algorithm

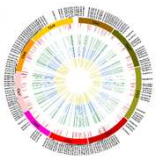


## Storage of molecular data

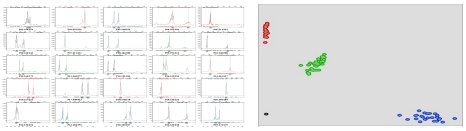
sequence data

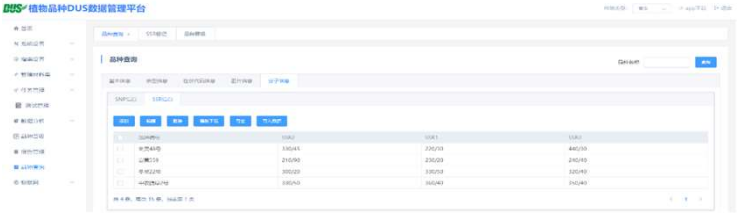


selection of marks

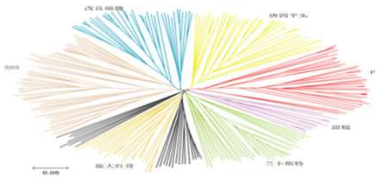


fingerprint image



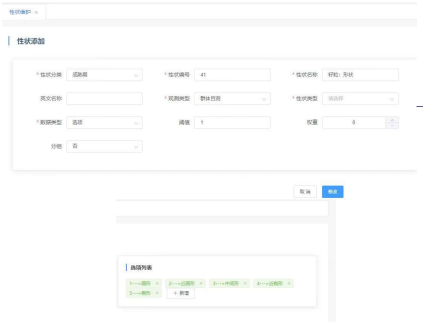


data management

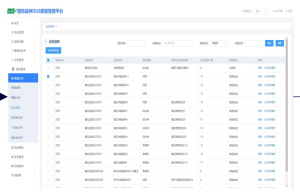


cluster analysis

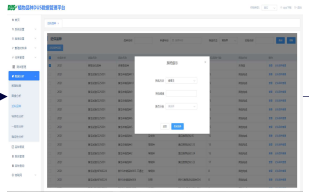
## Procedures of distinctness analysis



Set up TG parameters

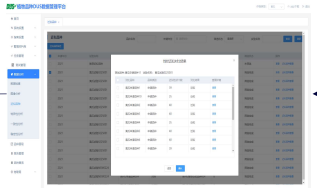


candidate varieties list



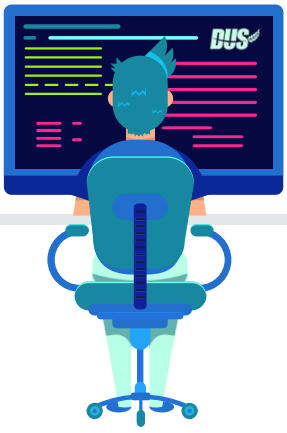
selection of similar varieties

report of distinctness by statistical analysis or side by side comparison



result of similar varieties

## 04 Plans



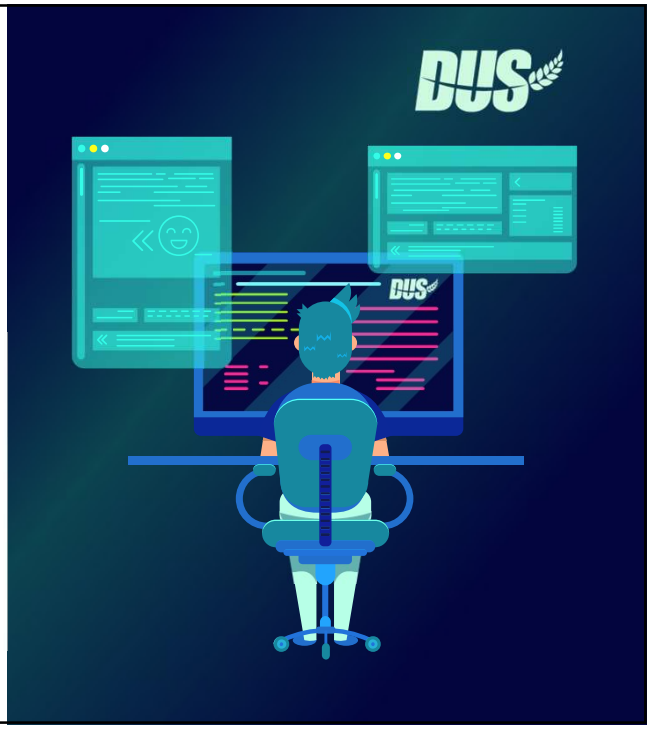
## Plans

**2023 v6.0**
forecast model between pheno-, geno-, enviro-data

**2024 v7.0**
automatic data record by unmanned vehicles

**2025 v8.0**
automatic test for all time and all space

# Thank you for your attention!

[End of Annex and of document]