

Generation Challenge Program

-

Introduction to GCP issues in data quality

Theo van Hintum

SP4 leader Generation Challenge Programme

Centre for Genetic Resources, The Netherlands (CGN)

Wageningen UR



Generation Challenge Programme

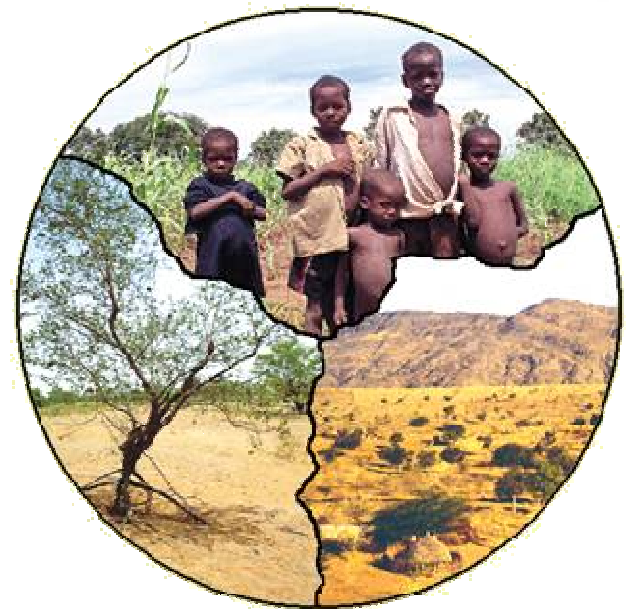
“I challenge the next generation to use new scientific tools and techniques to address the problems that plague the world’s poor”

Dr. Norman Borlaug



Generation Challenge Programme

- elements of presentation
 - brief introduction to GCP objectives and approaches
 - data quality issues in GCP

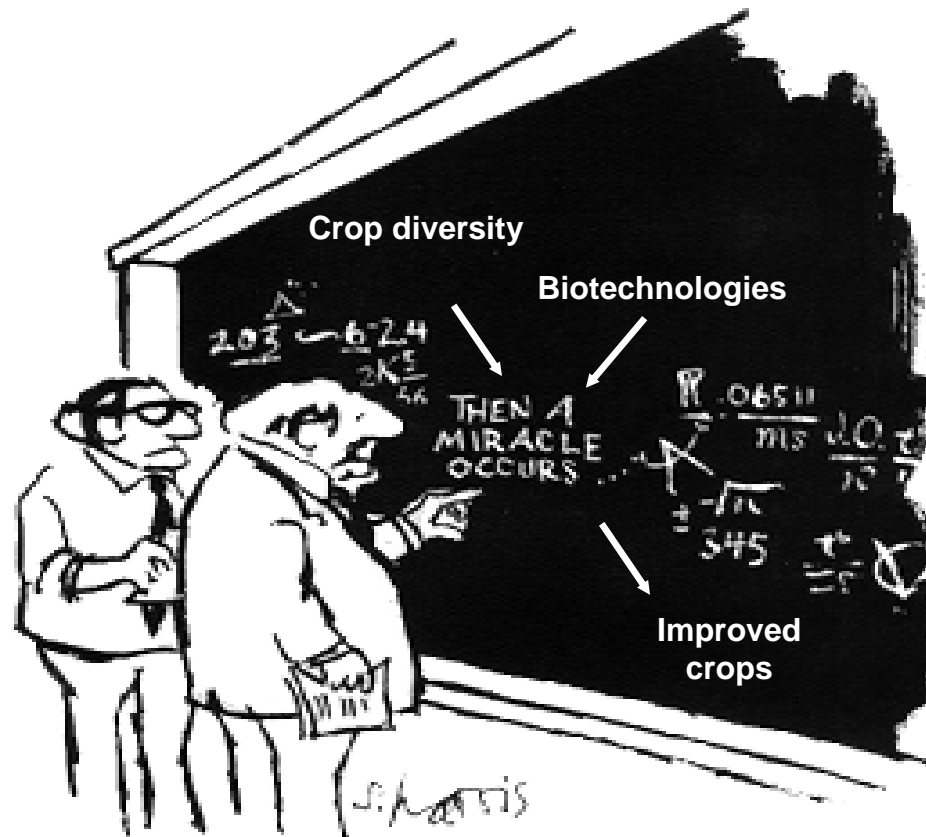


Generation Challenge Programme

- mission
 - use advanced genomics science and plant genetic diversity to overcome complex agricultural bottlenecks that condemn millions of the world's neediest people to a future of poverty and hunger



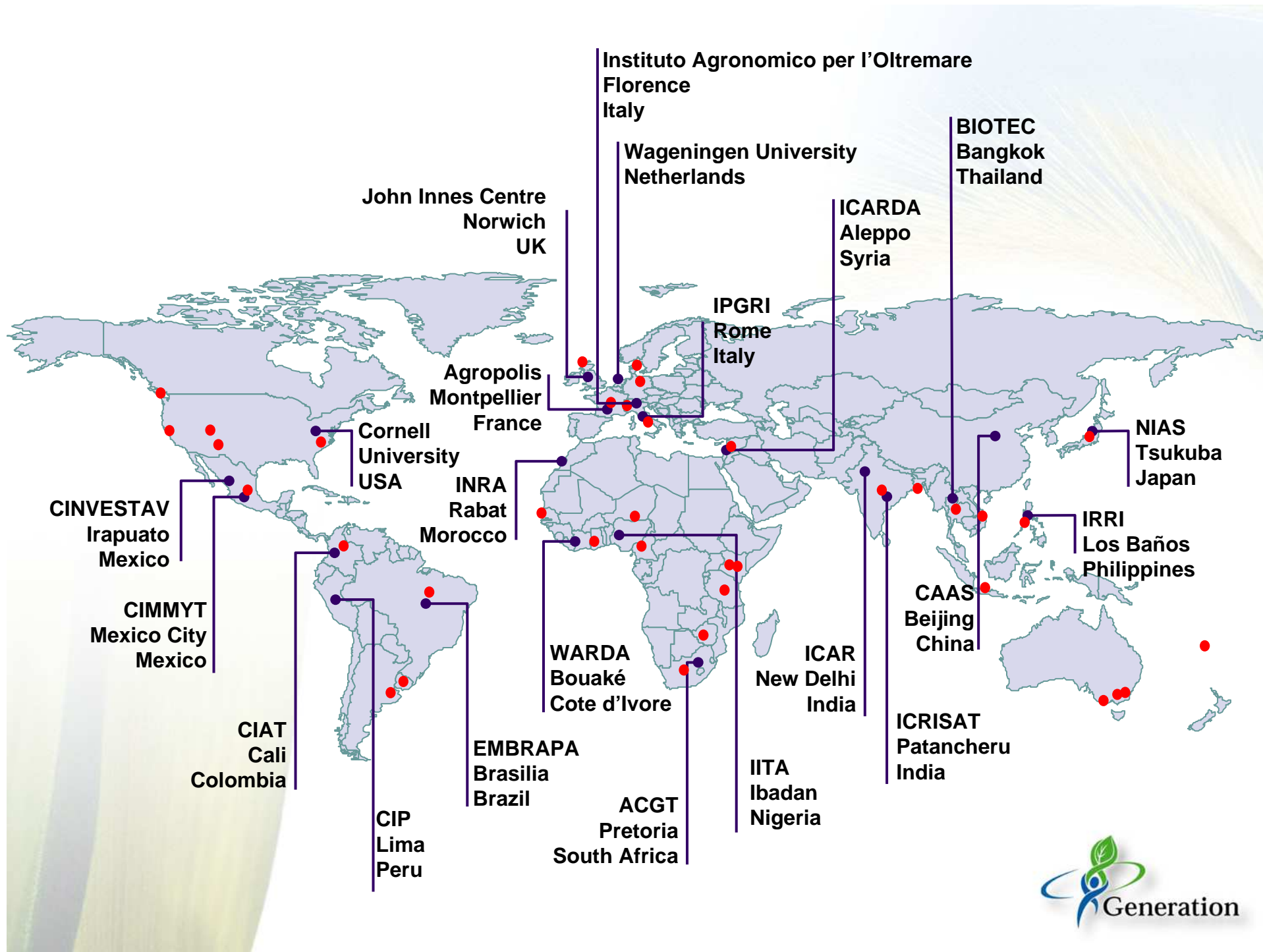
Generation Challenge Programme



"I think you should be more explicit here in step two."

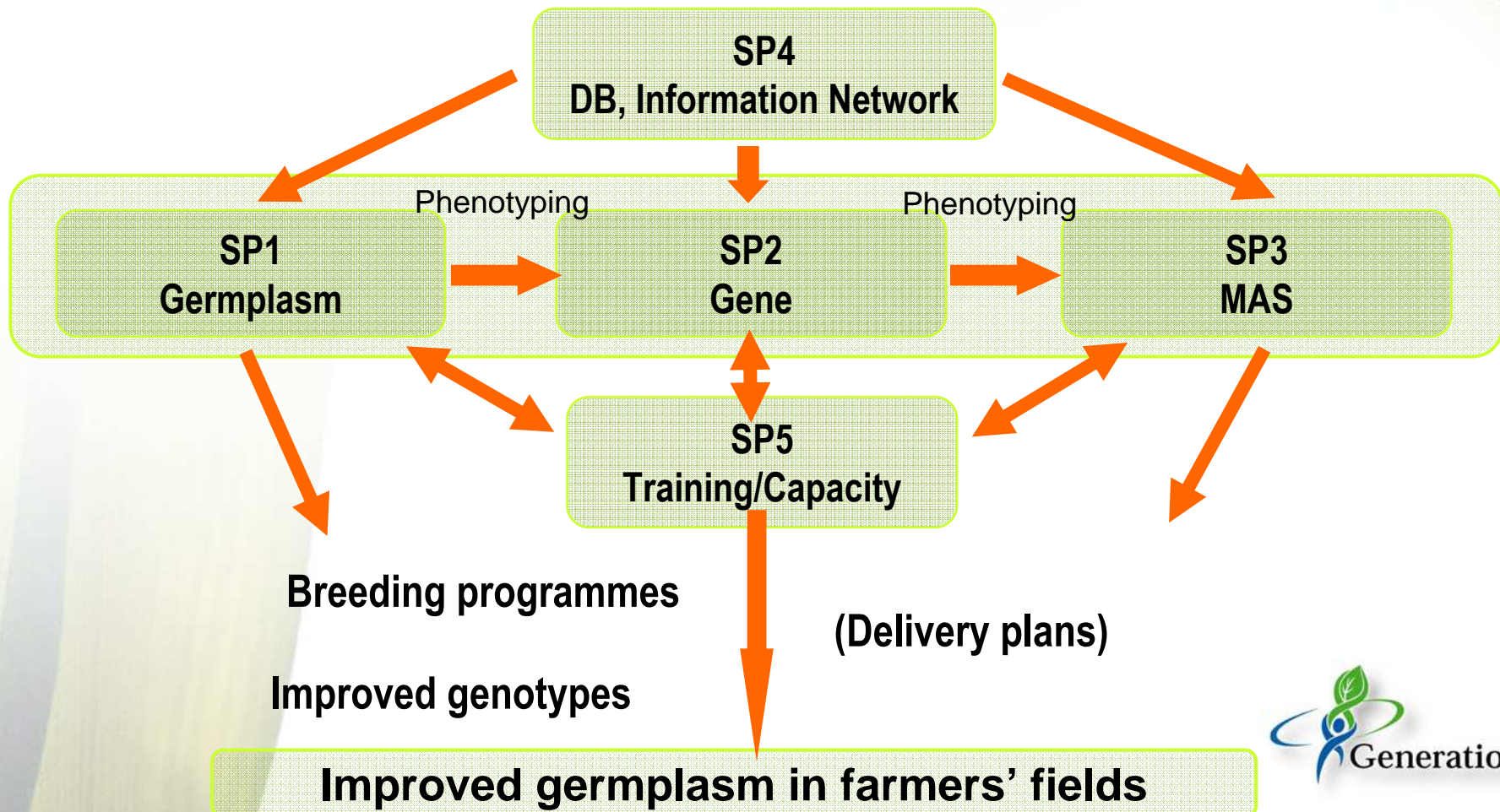
Generation Challenge Programme

- the actors
 - GCP is a multidisciplinary ‘true’ collaboration in the plant sciences
 - ‘Advanced Research Institutes’
 - CGIAR institutes
 - ‘National Agricultural Research Stations’



Generation Challenge Programme

- research structure



Generation Challenge Programme

- additional facts
 - launched in August 2003
 - 10 year framework two 5 year periods
 - target areas
 - marginal environments in Sub-Saharan Africa, Latin America and South East Asia
 - mandate crops: all CGIAR crops
 - budget c. \$16M / year
 - major donors
 - European Union
 - Bill and Melinda Gates Foundation
 - DFID
 - World Bank





GCP Data Quality Issues



GCP Data Quality Issues

- data quality is **hot** in the CGP
 - issue in commissioned resources building activities
 - in hypothesis driven (competitive) research the data produced are used in the project and are generally of better quality
 - here data availability is an issue

GCP Data Quality Issues

- GCP made a flying start
 - data production started without a proper strategy
 - large scale genotyping of germplasm collections
 - genotyping done in small scale labs
 - no provisions for data quality or availability
 - no methodology available
 - no provisions in the contracts
- obviously it became clear that data quality and availability was an issue

GCP Data Quality Issues

- data availability is tackled in several ways
 - GCP Central Registry plus data templates
 - increasing application of web services
 - standard clause in contracts about availability
- data quality remains a problem
 - genotyping data quality is being determined by repeating
 - quality indicators are being designed
 - data quality is multi-dimensional property
 - completeness / consistency / correctness
 - for quality report in GCP Central Registry

GCP Data Quality Issues

- GCP actions to improve quality of data
 - increase visibility of data
 - application of quality indicators
 - does it comply to format
 - Data Resolution
 - will result in better curation
 - workshops on aspects of data quality
 - to raise awareness
 - showed how bad the situation was in some cases
 - out-source genotyping to quality labs
 - support quality management
 - support development of LIMS

