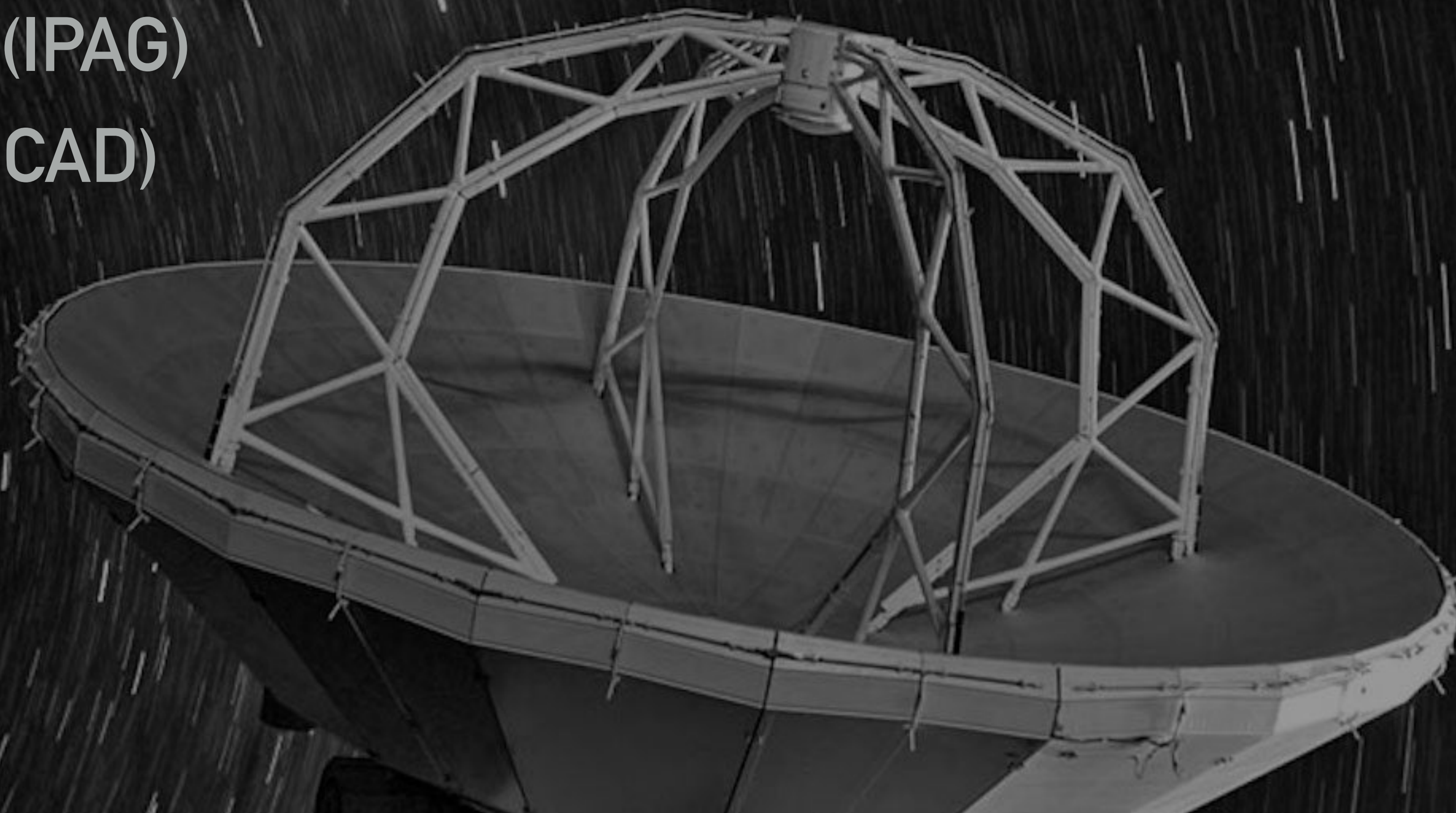


# TOWARDS REPRODUCIBLE RESEARCH WITH NIX: ASTRONOMICAL DATA REDUCTION ON GRICAD CLUSTERS

SÉBASTIEN MARET (IPAG)

BRUNO BZEZIK (GRICAD)



## WHAT IS REPRODUCIBLE RESEARCH?

« A RESEARCH WORK IS CALLED REPRODUCIBLE IF ALL INFORMATION RELEVANT TO THE WORK, INCLUDING, BUT NOT LIMITED TO, TEXT, DATA AND CODE, IS MADE AVAILABLE, SUCH THAT AN INDEPENDENT RESEARCHER CAN REPRODUCE THE RESULTS »

Vandewalle et al. (2009), cited in « Vers une recherche reproductible », Desquilbet et al. (2019)  
<https://rr-france.github.io/bookrr/>

## WHY RESEARCH WORK NEED TO BE REPRODUCIBLE?

- ▶ Recent example: the claimed detection of phosphine (a possible biomarker) in the atmosphere of Venus (Greaves et al. 2020), re-analyzed by Snellen et al. a few days after.

The image shows a screenshot of a preprint article page. At the top left, the word "nature" is partially visible. A blue bar at the top right contains the word "ARTICLES". Below this, a white box contains the text "Astronomy & Astrophysics manuscript no. alma\_venus" and "October 21, 2020". On the right side of this box, it says "©ESO 2020". The main title of the article is "Re-analysis of the 267-GHz ALMA observations of Venus" followed by the subtitle "No statistically significant detection of phosphine★". The authors listed are "I.A.G. Snellen<sup>1</sup> L. Guzman-Ramirez<sup>1</sup> M.R. Hogerheijde<sup>1,2</sup> A.P.S. Hygate<sup>1</sup> F.F.S. van der Tak<sup>3,4</sup>". At the bottom, there are four footnotes providing the affiliations for each author: <sup>1</sup>Leiden Observatory, Leiden University, Postbus 9513, 2300 RA Leiden, The Netherlands; <sup>2</sup>Anton Pannekoek Institute for Astronomy, University of Amsterdam, Science Park 904, 1090 GE Amsterdam, The Netherlands; <sup>3</sup>SRON Netherlands Institute for Space Research, Landleven 12, 9747 AD Groningen, The Netherlands; and <sup>4</sup>Kapteyn Astronomical Institute, University of Groningen, Landleven 12, 9747 AD Groningen, The Netherlands. On the left side of the page, there is a vertical list of partial words: "P", "Jan", "Da", "Em", "Zhu", "Mea", "nea", "phir", "spe", "PH<sub>3</sub>", and "ph".

nature

ARTICLES

Astronomy & Astrophysics manuscript no. alma\_venus  
October 21, 2020

©ESO 2020

**Re-analysis of the 267-GHz ALMA observations of Venus**  
**No statistically significant detection of phosphine★**

I.A.G. Snellen<sup>1</sup> L. Guzman-Ramirez<sup>1</sup> M.R. Hogerheijde<sup>1,2</sup> A.P.S. Hygate<sup>1</sup> F.F.S. van der Tak<sup>3,4</sup>

<sup>1</sup>Leiden Observatory, Leiden University, Postbus 9513, 2300 RA Leiden, The Netherlands  
<sup>2</sup>Anton Pannekoek Institute for Astronomy, University of Amsterdam, Science Park 904, 1090 GE Amsterdam, The Netherlands  
<sup>3</sup>SRON Netherlands Institute for Space Research, Landleven 12, 9747 AD Groningen, The Netherlands  
<sup>4</sup>Kapteyn Astronomical Institute, University of Groningen, Landleven 12, 9747 AD Groningen, The Netherlands

P  
Jan  
Da  
Em  
Zhu  
Mea  
nea  
phir  
spe  
PH<sub>3</sub>  
ph

## WHAT IS NIX AND HOW IT CAN HELP WITH REPRODUCIBILITY?

- ▶ **Nix** is a package manager for Linux and other Unix systems (including macOS). It comes with a collection over 60,000 packages called **Nixpkgs**.
- ▶ Nix makes package management:
  - ▶ **Reproducible:** packages are built in isolation from each other, so there are no hidden dependencies.
  - ▶ **Declarative:** environments for software projects (in any programming language: Python, C...) can be build quite easily.
  - ▶ **Reliable:** installing or upgrading one package cannot break other packages.

## WHAT IS NIX AND HOW IT CAN HELP WITH REPRODUCIBILITY?

- ▶ In practice, Nix can be used:
  - ▶ to install software packages in the user's profile, such as a Python interpreter, a C library, etc, using the `nix-env` command.
  - ▶ to build **isolated software environments** for a project using the `nix-shell` command.
- ▶ On Luke and Dahu, nix shells can be build on any node, not only the frontend. This allows to create isolated environnement within an OAR job.

### EXAMPLE: REDUCTION OF ALMA DATA ON ON GRICAD CLUSTERS

- ▶ ALMA is an interferometer located in the Atacama desert, in the Chilean Andes. It's the largest astronomical observatory.
- ▶ Observations are made public after one year in an archive. Typical datasets are up to a Tb.
- ▶ Data are calibrated, reduced and analysed on Luke. Jobs are sequential and takes up to a few hours.
- ▶ Several software packages: CASA, Gildas, and Python.



## CONCLUSIONS: THE GOOD, THE BAD AND THE UGLY

### THE GOOD

- ▶ Nix allows to create reproducible software environments for research projects.
- ▶ Environnement can be shared with GitHub and archived on Zenodo (with a DOI)

### THE BAD

- ▶ Nix is not user friendly.
- ▶ Pinning versions is not straightforward (this will improve with Nix Flakes).

### THE UGLY

- ▶ Writing Nix expression can be tricky and frustrating.