Towards a FAIR understanding of compact group evolution: A case study of HCG16

Image credit:
Jane Charlton (Penn. St. U.),
NASA, ESA, ESO

We separate the neutral gas (HI) content of HCG 16 into the gas associated with galaxies and extended features, revealing that although globally the group has not lost HI gas, galaxies A and B (right pair above) appear to have unbound much of their gas reservoir and the star formation events in galaxies C and D (left) are rapidly consuming gas and disturbing its distribution. This work was performed in an end-to-end open and reproducible manner, following “FAIR” principles. A complete workflow, built with docker containers and Conda environments, accompanies the paper. In addition, the final figures and analysis can be re-generated in cloud-based Jupyter notebooks.
(Jones et al 2019, A&A 632, 78)

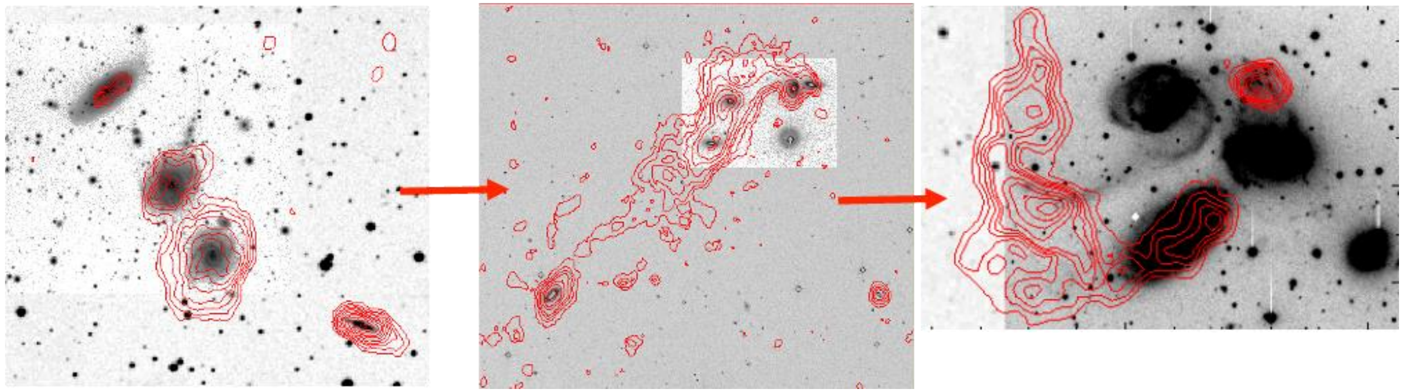
Mike Jones

Collaborators: L. Verdes-Montenegro, A. Damas, S. Borthakur, M. Yun, A. del Olmo, J. Perea, B. Williams, D. Lopez Gutierrez, F. Vogt, S. Luna, J. Garrido, S. Sanchez, J. Cannon & P. Ramírez

Context: HI morphology of compact groups and a reproducibly crisis in science

- Compact groups are **dense**, but **isolated**, **groups** of 4-10 galaxies with number densities comparable to cluster cores.
- They are generally found to be somewhat **deficient in HI** (neutral gas) and their HI morphologies appear to follow an evolutionary sequence (Verdes-Montenegro et al 2001):
  1. HI predominantly contained within galaxy discs.
  2. Significant quantities of HI in both galaxy disc and the intra-group medium.
  3. Galaxies almost devoid of HI gas.
- **HCG 16** is a prototypical **phase 2** compact group with many ongoing processes, which are expected to drive its transformation into phase 3.
- Science in general (including astronomy) is in the midst of a **reproducibility crisis** where many results in the literature cannot be reproduced, and in some cases researchers cannot even reproduce their own previous results.
- FAIR (which stands for Findable, Accessible, Interoperable and Reusable) is a bottom-up initiative (www.go-fair.org) to maximise the reusability of scientific data.
  - We used our analysis of HCG 16 as a real test case for implementing these principles in practice and even going beyond them by applying a similar approach to the software and methodology of our analysis, as well as the for the data themselves.

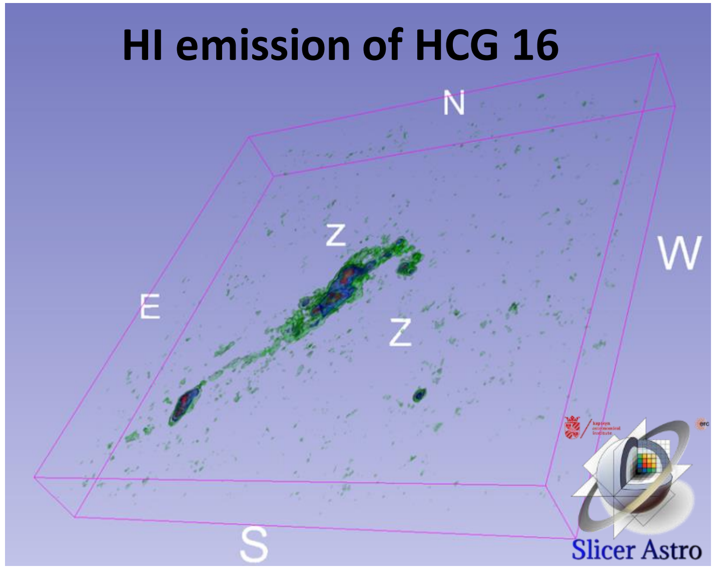
Phase I: Gas mostly within galaxies. **Phase II:** Clear tidal features appear. **Phase III:** Little gas remains in galaxies.



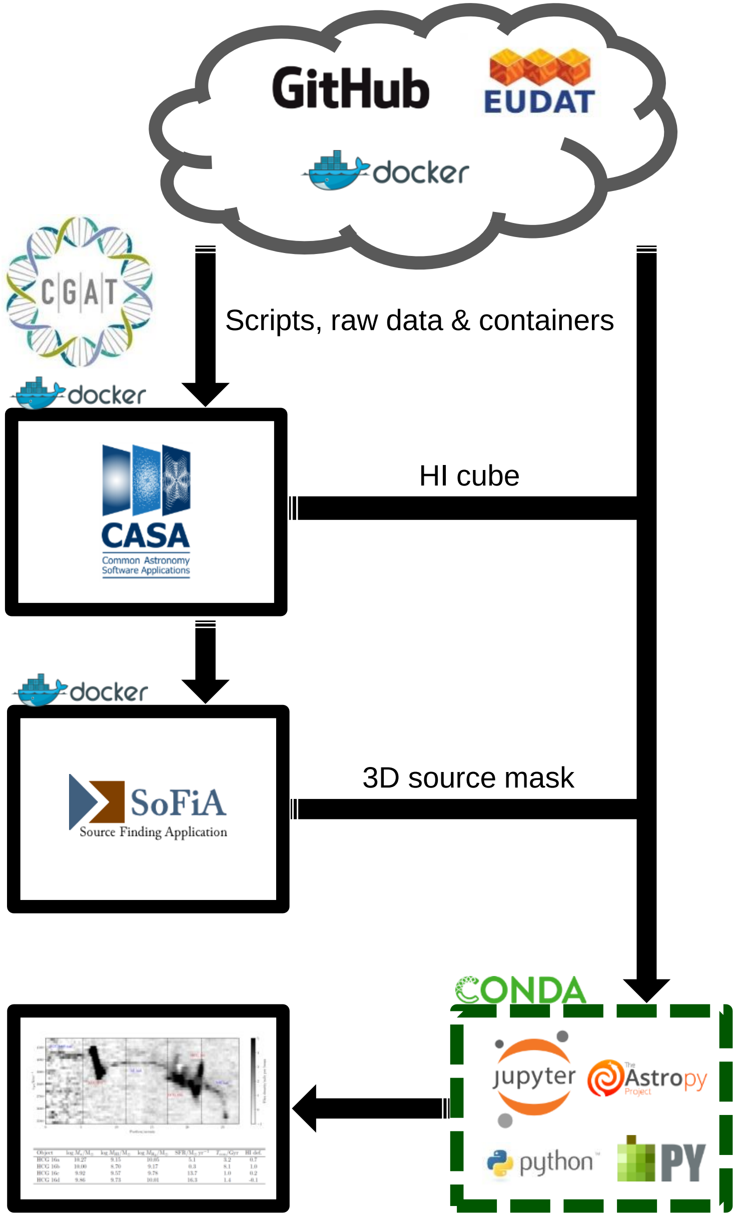
Methodology: Latest imaging tools and a containerised workflow

- HI emission in the group was imaged using the multi-scale CLEAN algorithm to robustly map extended features.
- The HI cube was inspected with the 3D visualisation tools SlicerAstro and X3D, and emission was separated into its galactic and extended components.
- HI-deficiency was estimated by comparison to the HI scaling relations of isolated galaxies (Jones et al 2018).

http://amiga.iaa.es/FCKeditor/UserFiles/X3D/HCG16/HCG16.html



- A complete workflow from raw data to final figures is preserved in GitHub (www.github.com/AMIGA-IAA/hcg-16).
- Software environments and scripts are controlled with docker containers and conda environments.
- CGAT-core, a workflow management system, encodes the steps of the workflow to streamline execution.
- The raw and final data are stored in EUDAT and the Jupyter notebooks to generate the paper's figures can be executed and modified in the cloud.



Results: HI content and morphology

- Despite being attached to an enormous HI tail, **NGC 848** has an almost **regular HI disc**, both in terms of mass and kinematics.

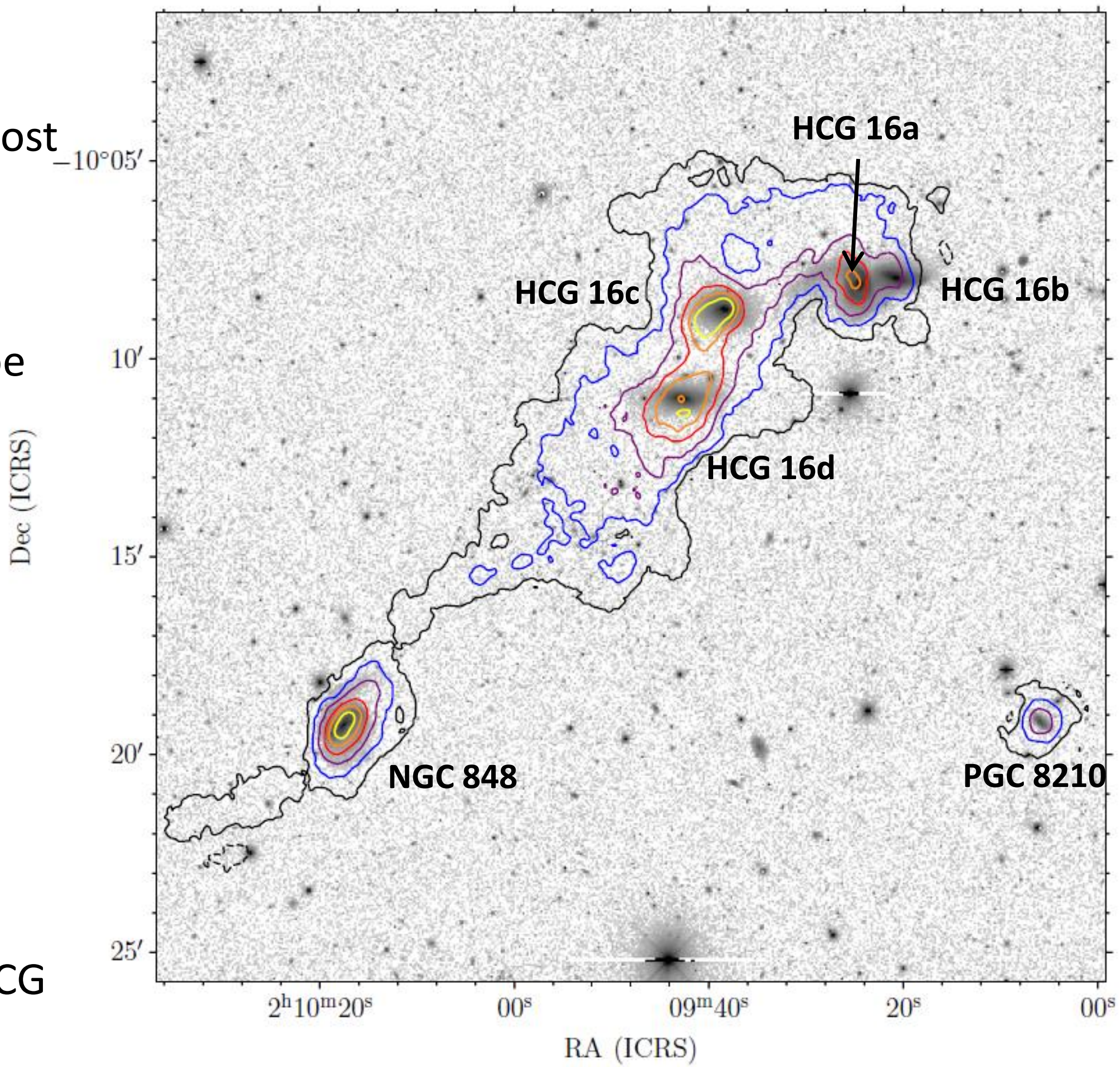
- The slight **velocity gradient** in the HI emission of HCG 16d is almost **perpendicular** to its stellar disc. The most likely cause is the outflow from its starburst event disrupting its HI disc.

- There are several dense clumps of HI over 10^8 M_{sol}, which may be the progenitors of **tidal dwarf galaxies**.

- The overall **HI mass** of HCG 16 is **consistent** with that expected given the optical luminosity of its constituent galaxies.

- However, only **HCG 16a and b** show **significant HI-deficiency**, strongly suggesting that the majority of the gas in extended features originated in those two galaxies.

- NGC 848 is connected to a continuous (both spatially and kinematically) stream that extends all the way to just north of HCG 16a. The passage of NGC 848 through the group likely stretched out already loosely bound gas around HCG 16a and b.



Results: Reproducing the HI analysis of HCG 16

- We were successful in creating a **full pipeline** to reproduce the exact numerical results from the paper. This can be **executed with a single command** on (almost) any unix-based system. The only external dependencies are bash and docker.

- The raw and final data are both stored in the EUDAT service (and have DOIs), and the final data are also stored in CDS.

- The software environments to perform the data reduction and analysis are maintained as docker containers (in dockerhub) and conda environments. In general **conda environments** are preferable as they are **easier to work with** and don't require root access, however, **docker containers** are more general and can be used to **preserve specific versions** of poorly maintained software.

- The reduction and analysis **scripts** are all stored **outside these containers/environments** such that any end user can modify parameters in the reduction and analysis steps and re-run the pipeline. This design enables **reusability** in addition to just being able to reproduce the exact results of the original paper.

- A few aspects of the paper are not contained in the pipeline, such as an analysis step reliant on **manual interaction** and an image enhancement performed by a collaborator. In such cases we focused on **preserving** what was done (i.e. the final output of these steps) to allow results based on these steps to be reproduced. However, these steps themselves are not included in the pipeline in a reusable way.

- Finally, in addition to the pipeline, the GitHub repository also hosts **Jupyter notebooks** to generate all the **final figures** are tables from the paper. These notebook can be executed and modified in the cloud-based service mybinder, without any need for local software dependencies, making it straightforward for anyone to reuse or repurpose these figures.

Conclusions

- The global HI content of HCG 16 is consistent with scaling relations of isolated galaxies, however, that gas is clearly highly disturbed.
- The most striking feature, the enormous tail connecting NGC 848 to the core group, likely formed as a result of a tidal interaction between the HCG 16ab pair and NGC 848 when it passed through the group (about 0.5 Gyr ago). This formed a continuous structure that can be traced in both position and velocity across the whole group (see figure).
- The star formation events in HCG 16 c and d are expected to rapidly consume much of their available fuel, while tidal processes will continue to unbind gas. This will results in a phase 3 group where little HI remains in the galaxies.
- The single most important lesson for reproducibility is that it's much easier if you start thinking about this at the start of a project than at the end.
- Providing Jupyter notebooks (if you make your figures in Python) in GitHub/mybinder to re-generate the final figures and tables of a paper is a simple, but significant, step towards aiding reproducibility and especially reusability.