What would be gained by implementing a DMP on the OIDB service already in production since 2015? Never too late for the DMP?

The aim of this poster is to highlight the gap between a formal data management plan and the practical approach applied to an already launched service . The analysis focuses on the main topics of the DMP and identifies how to solve or improve main issues given the help of a data warehouse under construction. The candidate project 'OiDB' operates since mid-2015 and provides a data portal dedicated to the optical interferometry domain. It collects observation logs or science ready files from several data repositories and facilities. The specifications of the project have been led by the associated scientific community in the Virtual Observatory framework. Many issues concerning the meta data have been discussed at the beginning of the project but no formal data management plan has been built. Let's try to complete a data management plan just now and see if it will point out any new issues. Are there any questions we should have asked before or any concerns we missed? The poster will report some feedback on the evolution of the data management process from a practical point of view. It will show if these evolutions could have been anticipated if we have had a DMP at the very beginning of the project. This analysis will help the enhancement of the Grenoble Observatory's datacenter (OSUG-DC) by pointing the priorities for new services such as data preservation or data identification.

Laurence Michaud [a], Bernard Boutherin [b], Guillaume Mella [ab]

[a] Univ. Grenoble Alpes, CNRS, IPAG, F-38000 Grenoble, France
[b] CNRS, OSUG, F-38000 Grenoble, France

DCC DMP	OIDB
1 – Administrative Data • Basic information e.g. project title, your name, contact details, reference numbers / Ids • A summary of the research to explain the purpose for which data are being collected • Details of related policies and procedures E.g. institutional data policy or departmental guidelines	*1 – Administrative Data* • Formal administrative data was not needed as the project was supported by the OLBIN collaborative community • OIDB is a portal to diffuse optical interferometric data. Each data has its own data PI. The data PI is the person in charge of the scientific program that has led the observation and the subsequent production of the calibrated data that will be diffused on the database • The 1st version of the data base was set up by its community with its own self-defined policies and procedures. ➢ **Using a DMP could have insured a better visibility to the data base describing its aims** ➢ **DMP points out that the identification of data PIs does not fulfil contact details. Improvement: a DMP for each source of data or the requirement of an unique identification**
2 - Data Collection • Are there any existing data that you can reuse? • What standards or methodologies will you use to create data? • Do your chosen formats and software enable sharing and long-term access to the data? • How ill you structure and name your folders and files? • How will you handle versioning? • What quality assurance processes will you adopt?	*2 - Data Collection* • The main part of the data are **public data** (regular harvesting of CDS, ESO and CHARA bases). • N/A astronomical observations and calibration • Data are stored in the **OIFITS format** (calibrated optical interferometric data) • The structure of the data base foresees no folder organisation. One repertory for each collection. • **Versioning is not yet supported**, but the published data is always the last version. • The data PIs are responsible for the accuracy and consistency of the data and metadata that will be diffused via the database service **. The possibility of guaranteeing the metadata via a quality flag is not used(less than 1%)** ➢ **Quality assurance processes could have been automated if foreseen** ➢ **Versioning has not been implemented though it has been discussed and finally postponed**
3 - Documentation and Metadata • Identification of Data • What documentation and metadata will accompany the data? What metadata standards will you use and why? • How will you capture / create this documentation and metadata?	*3 - Documentation and Metadata* • **DOI is not yet used** to identify data • The **ObsCore data model** developed by the IVOA describes precisely astronomical data and guarantees compatibility with other VO tools. It was extended for the needs of OIDB database. Metadata such as the number of spectral channel, visibilities, squared visibilities, instrument mode, telescope configuration have been added • OIFITS files are splitted into relevant **granule of information** (1 granule=1 target/1 night / 1 instrument mode / 1 OIFITS. The metadata of a granule often represents a subset of an OIFITS metadata.) ➢ **Using a DOI to identify data is foressen in OIDB 2.0 thanks OSUG Data Center services**
4 –Ethics and legal Compliance • Have you gained consent for data sharing and preservation? • How will you protect the identity of participants if required? E.g. via anonymisation? • Will data sharing be postponed / restricted? E.g. to publish or seek patents • How will the data be licensed for reuse?	*4 –Ethics and legal Compliance* • Data are partially public data or data PIs are voluntarily sharing their data → consent is underwritten. • N/A No need to protect the identity of participants. • Access to some data are restricted. The database automatically releases data access at the end of the embargo period. • The portal and all its content including all sort of data are protected under the **CC License** // Attribution-non commercial-Share Alike 4.0 International (CC BY NC SA 4.0) ➢ **DMP compliant**
5 - Storage and backup • Do you have sufficient storage or will you need to include charges for additional services? • Who will be responsible for backup and recovery? • What are the risks to data security and how will these be managed? • How will you ensure that collaborators can access your data securely?	*5 - Storage and backup* • JMMC is a part of **OSUG Data Center** which provides IT resources for OSUG's Observatory services: storage, backup and recovery. • Access to restricted data is **secured** through login and password • Risks is more mishandling than malicious act • Web servers respond to the secured "https protocol" ➢ **DMP compliant**
6 –Selection and Preservation • Which data must be retained or destroyed for contractual, legal, or regulatory purposes? • What are the foreseeable research uses for your data? • Which data should be preserved and potentially shared? • What is the long-term preservation plan for the dataset? • Have you costed in the time and effort required to prepare the data for preservation and sharing?	*6 –Selection and Preservation* • No destruction of data, on the contrary • Need to capitalise the efforts made to obtain science ready data (calibrated data) • Archives are extremely valuable to understand the observed objects on a **long temporal baseline**. • Internal resources ➢ **DMP compliant**
7 - Data Sharing • With whom will you share the data, and under what conditions? • When will you make the data available? • How will potential users find out about your data?	*7 - Data Sharing* • Data are opened to the interferometry research community (web interfaces and Virtual Observatory protocols). Make them accessible for a larger community is part of the fundamental aims of the database. • One of the main problem remains sharing one's own data. There are little volunteer data sharing. • Foreseen: to share quicklook plots and to connect the database to the registry **Virtual Observatory Applications for Astronomers** of the *International Virtual Observatory Alliance* *http://ivoa.net/astronomers/applications.html* to enlarge visibility. ➢ **Visibility will be enlarged through registration to IVOA and through OSUG Data Center common practices across multiple services..**
8 – Responsibilities and Resources • Who is responsible for implementing the DMP, and ensuring it is reviewed and revised? What resources will you require to deliver your plan? Is additional specialist expertise or equipment required? • How will responsibilities be split across partner sites in collaborative research projects?	*8 – Responsibilities and Resources* • Internal project resources, benefit from human and technical resources (librarian colleague, OSUG Data Center) • External point of view, year basis meeting ➢ **DMP compliant**

References:

- *OIDB homepage http://oidb.jmmc.fr*
- *X. Haubois & al. A Global Database for Optical Interferometry SPIE Vol.9146 91460O-1 doi: 10.1117/12.2056977*
- *IVOA* http://ivoa.net/astronomers/vo_glossary.html
- *DCC. (2013) Checklist for a Data Management Plan v.4.0. Edinburgh: Digital Curation Centre. http://www.dcc.ac.uk/resources/data-management-plans*
- *OSUG Data Center http://datacenter.osug.fr/*

Thanks to this exercise around the creation of a data management plan, we will improve the **formalism** of the OIDB project summarising in a single document information that are split partly in the web site, partly in publications, partly in the heads of the project leaders!

The most important improvement pointed out remains digital **identification** of both **PIs** and **data**. Most of OIDB data are harvested from other databases, it though directly depends on the applied standards of these data sources.

All the questions around data, metadata and security are **well documented** and are **DMP compliant**.

It's **not to late for our DMP** as a 2.0 improved version of the portal will soon be accessible to public. And the services provided by **OSUG Data Center** will enable a better identification of data and a better visibility of the data base.

Library and Information Services in Astronomy (LISA) VIII,
June 2017, Strasbourg (France)