



# Open Data @ INAF

Cristina Knapic

# Summary

- Context;
  - what we mean for open data;
  - which kind of data I will talk about;
  - golden rules and recommendations;
- Public data archives;
  - adherence of archives and related services to Open Data Recommendations;
  - data policies;
- Open Data repository;
- DOIs;
- How to improve the Openness

# Context : Open Data in Italy

## Definizioni:

“I dati della ricerca sono definiti dal Decreto come documenti informatici, diversi dalle pubblicazioni scientifiche, raccolti o prodotti nel corso della ricerca scientifica e utilizzati come elementi di prova nel processo di ricerca, o comunemente accettati nella comunità di ricerca come necessari per convalidare le conclusioni e i risultati della ricerca.

I dati da considerare, come da definizione, sono quelli raccolti o prodotti nel corso della ricerca scientifica finanziata con fondi pubblici, utilizzati come elementi di prova nel processo di ricerca o necessari per convalidare le conclusioni e i risultati della ricerca, e resi pubblici, anche attraverso l'archiviazione in una banca dati pubblica gestita a livello istituzionale o su base tematica, da ricercatori, organizzazioni che svolgono attività di ricerca e organizzazioni che finanziano la ricerca.

Nel processo di identificazione dei dati della ricerca, da rendere disponibili per il riutilizzo, è necessario tenere conto della normativa in materia di protezione dei dati personali, degli interessi commerciali, dei diritti di proprietà intellettuale e dei diritti di proprietà industriale.

Inoltre, i dati della ricerca DEVONO rispettare i requisiti di reperibilità, accessibilità, interoperabilità e riutilizzabilità che rappresentano i 4 principi del framework FAIR (Findable - Accessible - Interoperable - Reusable)”

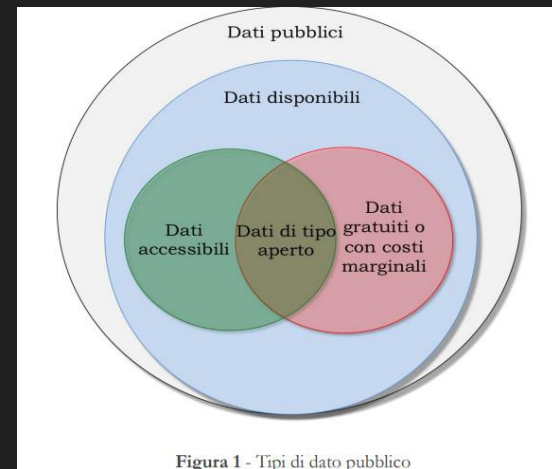


Figura 1 - Tipi di dato pubblico

# Context : Data under examination

- Archives and related services of astrophysical products
  - IA2 archives of national and international telescopes;
  - hosted archives;
  - simulations archives and repositories;
  - IA2 VOSpace;
- Open Access
  - Data sets;
  - Catalogues;
  - software;
- INAF DOIs
  - datasets;
  - software;

Formato	Non-proprietario	Leggibile meccanicamente	Stelle raggiungibili
RDF	Si	Si	★★★★
XML	Si	Si	★★★
JSON	Si	Si	★★★
CSV	Si	Si	★★★
ODS	Si	Prevalentemente	★★★
XLSX	Si	Prevalentemente	★★★
XLS	No	Prevalentemente	★★
TXT	Si	Prevalentemente	★*
HTML	Si	Prevalentemente	★*
PDF	Si	No	★
DOCX	Si	No	★
ODT	Si	No	★
PNG	Si	No	★
GIF	No	No	★
JPG/JPEG	No	No	★
TIFF	No	No	★
DOC	No	No	★

Tabella 1 - Formati più comuni per i dati aperti e relativi livelli di apertura

# Context : Golden roles and recommendations

## 1. Open Works

An open **work** *must* satisfy the following requirements in its distribution:

- **OPEN LICENSE OR STATUS** : The **work** *must* be in the **public domain** or provided under an open **license** (as defined in Section 2). Any additional terms accompanying the work (such as a terms of use, or patents held by the licensor) *must not* contradict the work's public domain status or terms of the license.
- **ACCESS**: The **work** *must* be provided as a whole and at no more than a reasonable one-time reproduction cost, and *should* be downloadable via the Internet without charge. Any additional information necessary for license compliance (such as names of contributors required for compliance with attribution requirements) *must* also accompany the work.
- **MACHINE READABILITY**: The **work** *must* be provided in a form readily processable by a computer and where the individual elements of the work can be easily accessed and modified.
- **OPEN FORMAT**: The **work** *must* be provided in an open format. An open format is one which places no restrictions, monetary or otherwise, upon its use and can be fully processed with at least one free/libre/open-source software tool.
- **Open Licenses**: A license *should* be compatible with other open licenses. A **license** is open if its terms satisfy the following conditions:
  - **REQUIRED PERMISSIONS**: The **license** *must* irrevocably permit (or allow) the following:
    - **USE**: The **license** *must* allow free use of the licensed work.
    - **REDISTRIBUTION**: The **license** *must* allow redistribution of the licensed work, including sale, whether on its own or as part of a collection made from works from different sources.
    - **MODIFICATION**: The **license** *must* allow the creation of derivatives of the licensed work and allow the distribution of such derivatives under the same terms of the original licensed work.
    - **SEPARATION**: The **license** *must* allow any part of the work to be freely used, distributed, or modified separately from any other part of the work or from any collection of works in which it was originally distributed. All parties who receive any distribution of any part of a work within the terms of the original license *should* have the same rights as those that are granted in conjunction with the original work.
    - **COMPILATION**: The **license** *must* allow the licensed work to be distributed along with other distinct works without placing restrictions on these other works.
    - **NON-DISCRIMINATION**: The **license** *must not* discriminate against any person or group.
    - **PROPAGATION**: The rights attached to the work *must* apply to all to whom it is redistributed without the need to agree to any additional legal terms.
    - **APPLICATION TO ANY PURPOSE**: The **license** *must* allow use, redistribution, modification, and compilation for any purpose. The license *must not* restrict anyone from making use of the work in a specific field of endeavor.
    - **NO CHARGE**: The **license** *must not* impose any fee arrangement, royalty, or other compensation or monetary remuneration as part of its conditions.
  - **ACCEPTABLE CONDITIONS**: The **license** *must not* limit, make uncertain, or otherwise diminish the permissions required in Section 2.1 except by the following allowable conditions:
    - **ATTRIBUTION**: The **license** *may* require distributions of the work to include attribution of contributors, rights holders, sponsors, and creators as long as any such prescriptions are not onerous.
    - **INTEGRITY**: The **license** *may* require that modified versions of a licensed work carry a different name or version number from the original work or otherwise indicate what changes have been made.
    - **SHARE-ALIKE**: The **license** *may* require distributions of the work to remain under the same license or a similar license.
    - **NOTICE**: The **license** *may* require retention of copyright notices and identification of the license.
    - **SOURCE**: The **license** *may* require that anyone distributing the work provide recipients with access to the preferred form for making modifications.
    - **TECHNICAL RESTRICTION PROHIBITION**: The **license** *may* require that distributions of the work remain free of any technical measures that would restrict the exercise of otherwise allowed rights.
    - **NON-AGGRESSION**: The **license** *may* require modifiers to grant the public additional permissions (for example, patent licenses) as required for exercise of the rights allowed by the license. The license may also condition permissions on not aggressing against licensees with respect to exercising any allowed right (again, for example, patent litigation).

# Public Data Archives in INAF

- Archives of INAF Observatory Facilities
- Archives of synthetic data produced by INAF researchers
- Hosted Archives
- Institutional data repositories: VOSpace, OwnCloud, Google Suite (?)

## Adherence:

- OPEN LICENSE OR STATUS
- ACCESS
- MACHINE READABILITY
- OPEN FORMAT
- Open Licenses
  - USE
  - REDISTRIBUTION
  - MODIFICATION
  - SEPARATION
  - COMPILATION
  - NON-DISCRIMINATION
  - PROPAGATION
  - APPLICATION TO ANY PURPOSE
  - NO CHARGE
  - ACCEPTABLE CONDITIONS
  - ATTRIBUTION
  - INTEGRITY
  - SHARE-ALIKE
  - NOTICE
  - SOURCE
  - TECHNICAL RESTRICTION PROHIBITION
  - NON-AGGRESSION

## Data in archives:

- policy on RAW data: typically free after a proprietary period of 1 y (signed by DS);
- on advanced products : NO spread policy defined!
- no link to publications (we are working on that!)
- no efforts spent on assignment of IDs to data sets except when published
- no diffuse use of DMPs
- no data models, flow etc... for each collection
- no linked details on pipelines, codes, infrastructures..

## Data in Repositories:

- no standard access (yes only for VOSpace)
- no policy
- no licences
- no format
- .... NOT FAIR

# Public data archives: data policies

As mentioned in INAF there was defined for raw data a data publication policy. Currently most of the archives are exposed via Interoperable Astrophysical standards (TAP, SSAP, SIAP...) for raw data.

Higher level products depend sometimes on private pipelines, workflows and are international collaboration with different policies etc.. so the scientific product is not always public but should be FAIR.

How much are them FAIR?

In INAF there is no spread diffusion of best practices on Data Management, interoperable standards, best practices and golden roles to publish FAIR data.

A common document on policies should be stated to stimulate the FAIR data production.

FAIR facilitate Open Data.

# Open Data Repository

In the INAF DSpace implementation, till now, beside the great attention given to Articles and papers in general, report also a discrete interest in saving data.

Software is also important to be FAIR with the dataset (origin and reconstruction related data).

## Naviga per Tipo

Ordinamento:  Risultati/pagina

Mostrati i risultati da 1 a 21 di 21

<a href="#">article</a>	7522
<a href="#">book</a>	22
<a href="#">book part</a>	84
<a href="#">book/conference editing</a>	3
<a href="#">book/proceedings editing</a>	38
<a href="#">conference abstract</a>	171
<a href="#">conference paper</a>	1255
<a href="#">conference poster</a>	23
<a href="#">database</a>	254
<a href="#">dataset</a>	2
<a href="#">doctoral thesis</a>	2
<a href="#">magazine article</a>	24
<a href="#">multimedia</a>	107
<a href="#">patent</a>	3
<a href="#">permanent exhibition</a>	1
<a href="#">research report</a>	916
<a href="#">software</a>	17
<a href="#">technical report</a>	2497
<a href="#">temporary exhibition</a>	6
<a href="#">website</a>	7
<a href="#">working paper</a>	21

Mostrati i risultati da 1 a 21 di 21



# Open Data Repository & DOIs

Open Data related products:

- catalogues : 254 but none handled by INAF
- datasets: 2 products saved in the repo

INAF Assigned DOIs:

- software : 25 between pipelines, control software etc...
- datasets : 18 high level products

Are those data OPEN? For sure!

Are those data useful? Potentially, since only who knows their topic can find them.

Are they useful to do better science? Absolutely but not widely spread.

# How to improve Openness ?

IF Open Data tends to find, understand and reuse data, we should promote

- Training on best practices of OS and OD in particular (data management, open formats, data models and flows);
- Participation to the World most representative associations of research for scientific data interoperability (IVOA, RDA, WDS, EOSC ...);
- Support existing teams on FAIRization of archives;
- Definition of an institutional policy on data;
- Increase the features and support the OA repository in linking scientific results, publications, and archives (i.e. TelBib);
- Increase the features of OA repository in adding scientific metadescrptors;
- Add to DOIs metadata astrophysical information
- ....

Thank you for your attention