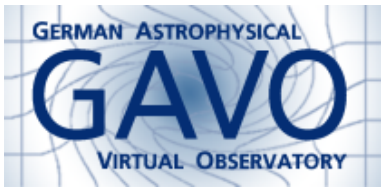
Publishing Scanned Plates Using DaCHS

Florian Rothmaier, Markus Demleitner, Joachim Krautter,
Holger Mandel, Stephanie Schwemmer, Otmar Stahl

Zentrum für Astronomie Heidelberg

Astroplate 2014, Prague, Vila Lanna

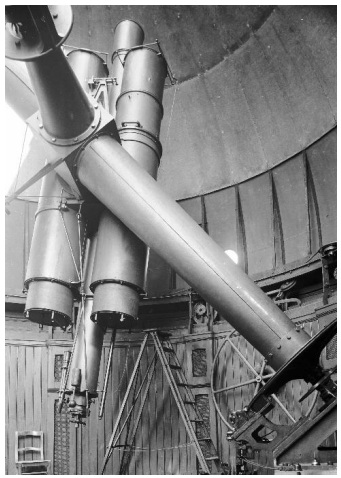


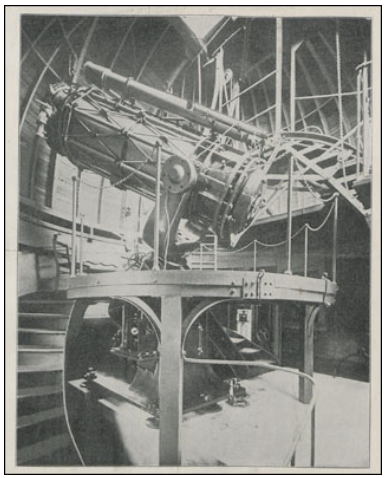
Outline

What Do We Mean by "Preservation of Photographic Plates"?

The Heidelberg Königstuhl Archives (I)

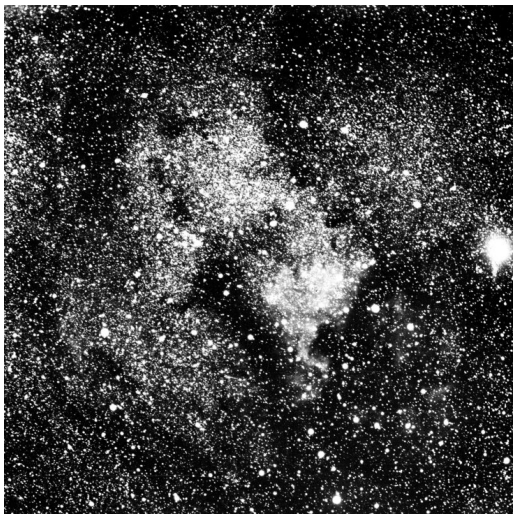
- about 25,000 astronomical photographic plates from various telescopes dating from the late 19th century to our days

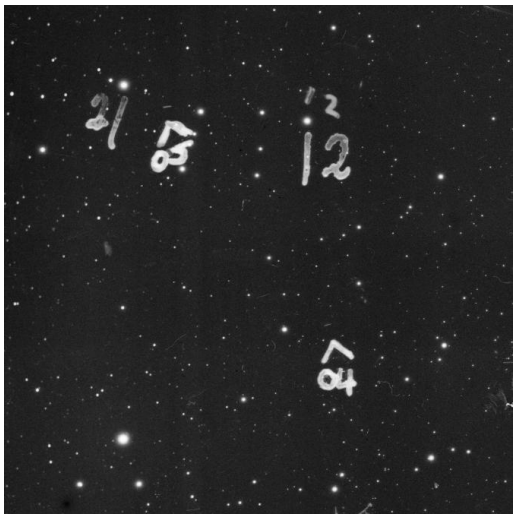





- plates were taken at various instruments
  - Bruce telescope at Königstuhl: $\sim$ 10,000 plates
  - Wolf's double astrograph at Heidelberg-Märzgasse/Königstuhl: $\sim$ 8,000 plates
  - Waltz reflector at Königstuhl: $\sim$ 5,000 photo plates
  - Schmidt telescope at Calar Alto: 400 plates

The Heidelberg Königstuhl Archives (II)





- covering a wide variety of objects of different types
  - nebulae: e.g. NGC 7000 (North America Nebula) taken with Wolf's Double Astrograph in three consecutive nights, September 11-13, 1891
  - minor planets: e.g. the asteroids (325) Heidelberga (label "05") and (175) Andromache (label "04"), January 17, 1909
  - comets: e.g. C/1911 O1 ("Brooks"), September 25, 1911

➥ see poster session for more details on the Königstuhl archives

Preservation of Photo Plates

- preservation of photo plates is a two-step process

Scanning Plates

- ensures the preservation of information despite ongoing oxidative deterioration of the photo emulsions

Publishing Plates

- ensures the persistence of the archives and the standardised access to them

Preservation of Photo Plates

- preservation of photo plates is a two-step process

Scanning Plates

- ensures the preservation of information despite ongoing oxidative deterioration of the photo emulsions

Publishing Plates

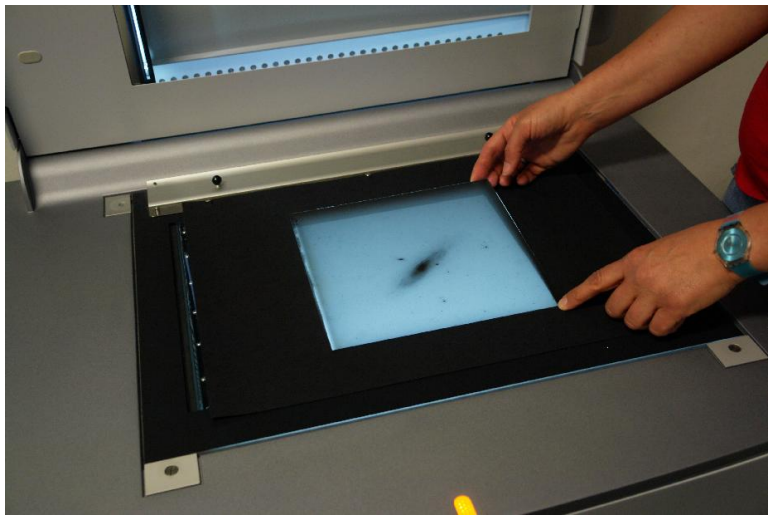
- ensures the persistence of the archives and the standardised access to them

- Heidelberg-Königstuhl scan project (2005-2014):
  - digitization of the Königstuhl archives
  - subsequent storage and publication of the scanned plates by using the infrastructure of the German Astrophysical Virtual Observatory (GAVO)

How Do We Digitise Our Photographic Plates?

Scanning the Plates

- Nexscan F4100 professional scanner of the "Heidelberg Druckmaschinen AG"
  - 315 mm x 457 mm scan area which is masked down to the various plate sizes
  - dynamic range: 16 bit/px
  - 5080 dpi optical resolution
  - chosen resolution for the scans: 2550 dpi (compromise between the requirements needed for a good scan and the amount of data per digitised plate)



Adding Metadata to the Plates

- digitised plates are stored in the commonly used FITS (Flexible Image Transport System) format with minimal headers (basically, NAXISn and BITPIX)
- observation journals are digitised into a plate database providing information about e.g. the observer, the observation time, and the exposure time
- information from the journals together with some plate-specific metadata (e.g. the instrument name or the photo emulsion used) is added to the FITS headers
- program SExtractor writes the objects found on a photo plate in (x, y) to a catalogue
- Astrometry.net tool is used for assigning $(x, y) \rightarrow (\alpha, \delta)$

Which Publication Infrastructure Is Available in the Virtual Observatory?

A Few Words on the Virtual Observatory

The Virtual Observatory (VO) is about...

Archiving Data

storage of all types of astronomical data, e.g.

- catalogues,
- images,
- spectra,
- ...

in data centres which can be accessed from all over the world

Developing Applications

interoperable tools that can be used to

- retrieve data taken by telescopes from all over the world
- analyse, manipulate or visualise astronomical data sets

Defining "Standards"

development of specifications on how

- data should look like ("data models")
- two or more machines should query and exchange data ("protocols")

Standards in the VO (I)

To define standards means to give an answer to questions like...

- In which format should I store my data?
- How should the model that describes my data look like?
- Which metadata should come along with a piece of data?
- Which "language" should be used for communication between two or more machines?

Standards in the VO (I)

To define standards means to give an answer to questions like...

- In which format should I store my data?
- How should the model that describes my data look like?
- Which metadata should come along with a piece of data?
- Which "language" should be used for communication between two or more machines?

➥ large variety of data models and protocols defined by the Virtual Observatory community over the years

International
Virtual
Observatory
Alliance

Simple Image Access Specification
Version 1.0

IVOA Recommendation 2009-11-16

Standards in the VO (II)

- "Simple Image Access Protocol" (SIAP):
  - allows for retrieving image data out of a variety of astronomical repositories by using a uniform interface
  - for a query based on a certain sky region the service returns a list of image candidates

Standards in the VO (II)

- "Simple Image Access Protocol" (SIAP):
  - allows for retrieving image data out of a variety of astronomical repositories by using a uniform interface
  - for a query based on a certain sky region the service returns a list of image candidates
- "Table Access Protocol" (TAP)
  - service protocol for accessing general table data

Standards in the VO (II)

- "Simple Image Access Protocol" (SIAP):
  - allows for retrieving image data out of a variety of astronomical repositories by using a uniform interface
  - for a query based on a certain sky region the service returns a list of image candidates
- "Table Access Protocol" (TAP)
  - service protocol for accessing general table data
- "ObsCore Data Model"
  - data model used to describe observational data
  - includes data elements like the name of the astronomical object observed, the exposure time, and the URL required to download the data

➥ combining the ObsCore model and TAP: "ObsTAP"

Standards in the VO (III)

- Why publishing your archive according to VO Standards?
  - data can be discovered by in-client standard interfaces (e.g. in TOPCAT)
  - data can be used with standard clients
  - data comes along with standard metadata helping later scientific exploitation

Standards in the VO (III)

- Why publishing your archive according to VO Standards?
  - data can be discovered by in-client standard interfaces (e.g. in TOPCAT)
  - data can be used with standard clients
  - data comes along with standard metadata helping later scientific exploitation
- However, ...
  - many people get scared by looking at the Virtual Observatory specifications
  - the technical obstacles seem to be larger than the benefits of using VO Standards

Standards in the VO (III)

- Why publishing your archive according to VO Standards?
  - data can be discovered by in-client standard interfaces (e.g. in TOPCAT)
  - data can be used with standard clients
  - data comes along with standard metadata helping later scientific exploitation
- However, ...
  - many people get scared by looking at the Virtual Observatory specifications
  - the technical obstacles seem to be larger than the benefits of using VO Standards

➥ GAVO offers a piece of software that tries to lower that barrier

GAVO's Data Center Helper Suite (DaCHS)

- DaCHS...
  - is a multi-protocol VO server implementing all major VO protocols
  - contains a variety of parsers for input data files, e.g. text files or FITS images
  - contains the Stan templating system of Python's web development framework Nevow allowing for the publication of HTML form-based services and documentation pages

GAVO's Data Center Helper Suite (DaCHS)

- DaCHS...
  - is a multi-protocol VO server implementing all major VO protocols
  - contains a variety of parsers for input data files, e.g. text files or FITS images
  - contains the Stan templating system of Python's web development framework Nevow allowing for the publication of HTML form-based services and documentation pages
- central concept of DaCHS' publication infrastructure: the Resource Descriptor (RD):
  - an XML file (typically, there is one RD per data collection)
  - contains all the metadata for a data archive
  - generates the services to be provided for the data

How To Publish Your Plate Archive to the Virtual Observatory?

Steps Towards a VO Publication: DaCHS Installation

- the preferred way to run DaCHS is on Debian or compatible systems on which it can be installed from GAVO's APT repository:
  - add `deb http://vo.ari.uni-heidelberg.de/debian stable main` to the file */etc/apt/sources.list*
  - update the package cache: `sudo apt-get update`
  - install the Debian package: `sudo apt-get install gavodachs-server`

Steps Towards a VO Publication: DaCHS Installation

- the preferred way to run DaCHS is on Debian or compatible systems on which it can be installed from GAVO's APT repository:
  - add `deb http://vo.ari.uni-heidelberg.de/debian stable main` to the file */etc/apt/sources.list*
  - update the package cache: `sudo apt-get update`
  - install the Debian package: `sudo apt-get install gavodachs-server`

➥ after doing these steps, DaCHS is on your machine ready for use...

Steps Towards a VO Publication: Writing an RD (I)

- find a detailed description on how to write a Resource Descriptor at `http://docs.g-vo.org/DaCHS/tutorial.pdf`
- start writing your RD by giving the `<resource>` root element of your xml file followed by some metadata:

```
<resource resdir="lswscans" schema="lsw">
    <meta name="creationDate">2007-11-10T12:00:00Z</meta>
    <meta name="description">Scans of plates obtained at Landessternwarte
        Heidelberg-Königstuhl, 1880 through 1999.</meta>
    <meta name="title">HDAP -- Heidelberg Digitized Astronomical
        Plates</meta>
    ...
</resource>
```

Steps Towards a VO Publication: Writing an RD (II)

- define your database table(s):

```
<resource resdir="lswscans" schema="lsw">
    ...
    <table id="plates" onDisk="True" adql="True" mixin="//siap#pgs">
        <column name="exposure" tablehead="Exp. time"
            unit="s" ucd="time.duration;obs.exposure"
            description="Effective exposure time" verbLevel="15"/>
        <column name="object"  type="text"
            tablehead="Targ. Obj." ucd="meta.id"
            description="Special object on plate" verbLevel="15"/>
        ...
    </table>
</resource>
```

  - `<column>` elements set the fields of the database table
  - `<mixin>` means: run internal code to endow the table with everything needed for a specific service (here: SIAP)

Steps Towards a VO Publication: Writing an RD (III)

- write your data ingestion component:

```
<resource resdir="lswscans" schema="lsw">
    ...
    <data id="import" updating="True">
        <sources recurse="True">
            <pattern>data/part1/*.fits</pattern>
            <pattern>data/part2/*.fits</pattern>
        </sources>
        <fitsProdGrammar qnd="True" id="impGrammar">
            ...
        </fitsProdGrammar>
        <make table="plates" rowmaker="make_plates"/>
    </data>
</resource>
```

  - the input data sets are declared within the `<sources>` elements
  - grammar chosen here returns FITS headers as dictionaries, i.e. a sequence of string-to-string mappings
  - `<rowmaker>` elements turn these mappings into proper database rows

Steps Towards a VO Publication: Writing an RD (IV)

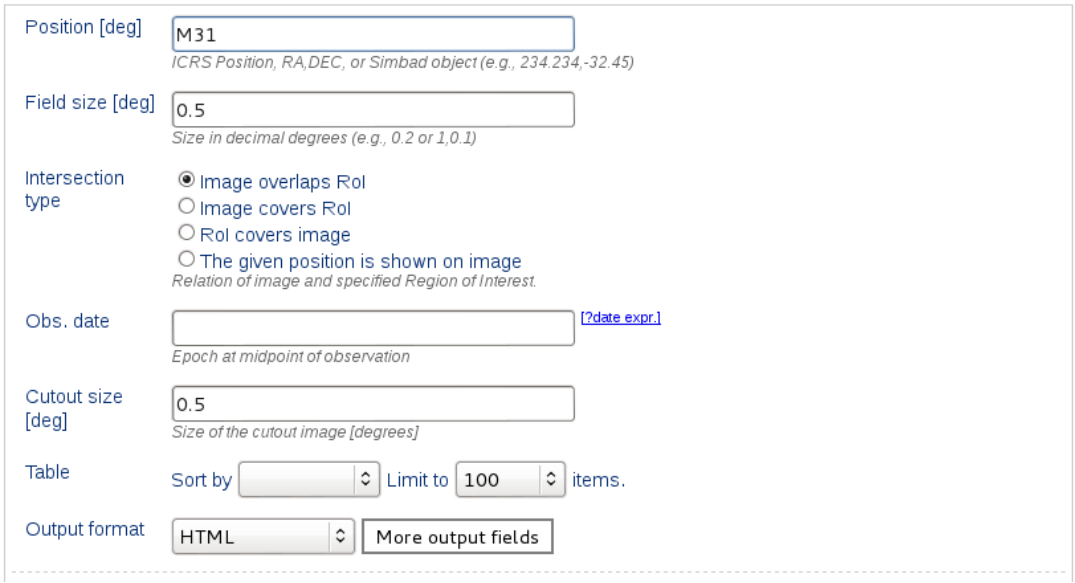
- define the service(s) exposing the data:

```
<resource resdir="lswscans" schema="lsw">
    ...
    <service id="siap" core="qhsiap" allowed="siap.xml">
        <publish render="siap.xml" sets="ivo_managed"/>
        <publish render="form" sets="ivo_managed" service="q"/>
        <meta name="shortName">hdap_siap</meta>
        ...
   </service>
</resource>
```

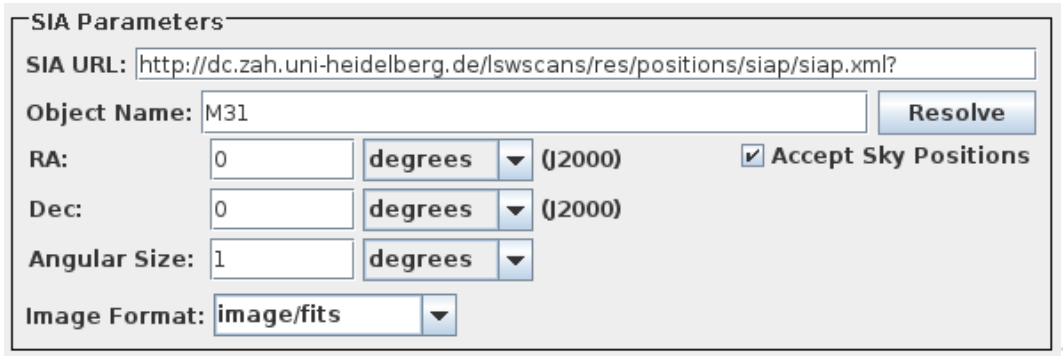
  - a service is a combination of a core and one or more renderers
  - a core is the element which performs the actual computations for the service
  - renderers set the interface(s), here a web form ("form") and an SIAP interface ("siap.xml")

And there It Is...

- `http://dc.zah.uni-heidelberg.de/lswscans/res/positions/q/form`



- TOPCAT's built-in SIAP client



Summary

- preservation of astronomical photographic plates is (at least) a two-step process comprising the digitisation and the publication of the data
- the software package DaCHS is in operation in various data centres all over the world
- due to its powerful publication infrastructure supporting to Virtual Observatory standards, DaCHS provides a reasonably smooth path for publishing scanned photo plates
- technical support for your data publication is available from the GAVO Heidelberg team at gavo@ari.uni-heidelberg.de

References

- The GAVO Data Center, http://dc.zah.uni-heidelberg.de/.
- Tody, D.; Plante, R. (2009), Simple Image Access Specification, IVOA Standards, http://www.ivoa.net/documents/SIA/20091116/.
- Dowler, P.; Rixon, G.; Tody, D. (2010), Table Access Protocol, IVOA Standards, http://www.ivoa.net/documents/TAP/.
- Louys, M. et al. (2011), Observation Data Model Core Components and its Implementation in the Table Access Protocol, IVOA Standards, http://www.ivoa.net/documents/ObsCore/index.html.
- HDAP – Heidelberg Digitized Astronomical Plates, http://dc.zah.uni-heidelberg.de/lswscans/res/positions/q/form.
- Full plate access to Heidelberg Digitized Astronomical Plates, http://dc.zah.uni-heidelberg.de/lswscans/res/positions/fullplates/form.