1 Publishable summary

1.1 Executive summary

Observational astronomy has dramatically evolved over the last 20 years as a result of the opening up of new observing windows in the electromagnetic spectrum and of the development of efficient and large area detectors attached to ground-based and space-borne instruments. With multi-wavelength measurements becoming available for billions of stars and galaxies, astronomy now ranks as a "big data" scientific activity and today many routine research activities rely on innovative data mining tools. However, large-scale multi-wavelength observational studies remain difficult to achieve for a large majority of astronomers. One of the most challenging steps of analysis is to associate measurements present in different catalogues and observed at different wavelengths to a unique astrophysical object and to assess the corresponding association probabilities. This problem is particularly acute in the case of cosmic X-ray sources detected by the ESA XMM-Newton satellite, which due to its specific high-energy optics, yields source positions of lower precision than available at many other wavelengths. In this context, ARCHES's main goal was to enrich the scientific usability of the catalogue of XMM-Newton sources by providing statistically qualified radio to X-ray spectral energy distributions for unresolved X-ray sources, and probabilistic identifications of extended X-ray sources with clusters of galaxies. In order to fulfil these objectives, careful assessment of the photometric and astrometric quality of multi-wavelength catalogues was first performed followed by the development i) of a forefront and so far unique multi-catalogue probabilistic cross-matching tool and ii) of an advanced integrated cluster finder using optical and near infrared imaging data to search for clusters of galaxies at any position in the sky. These tools were then applied to an enhanced and cleaned version of the XMM-Newton source catalogue to build over 200,000 spectral energy distributions of unresolved X-ray sources and over 500 clusters of galaxies candidates. Seven demonstration science cases focused on clusters of galaxies, active galactic nuclei and stellar activity highlighted the usability of ARCHES's tools and products and, most importantly, provided important feedback to the development of the tools and the cross-matched catalogues. ARCHES tools and products were presented to the international community in a workshop held in Paris. They are now accessible through a variety of interfaces specific to the project as well as through mainstream dissemination systems such as the Centre de Données astronomiques de Strasbourg.

1.2 Summary description of project context and objectives

The last decades have witnessed a tremendous increase of the observational astronomical data flow as a result of the development of new efficient and large area detectors supplemented with a computing power so far following Moore's law.

Ground-based and space-borne all-sky surveys have provided an essential astrometric and photometric reference frame and the first true maps of the entire sky.

This ever-increasing wealth of multi-wavelength data has fundamentally changed the way astronomers now tackle scientific problems. The previous paradigm, typically focusing on a single wavelength range, has now evolved towards a systematic fully multi-wavelength approach. In fact, our understanding of the physics of stars and galaxies now essentially rests on the modelling of their electromagnetic spectra over the widest range of frequencies, spanning from radio to the highest energy gamma-rays.

For instance, the number of catalogue entries in the Vizier service at the Centre de Données astronomiques de Strasbourg (CDS) which was about 500 million in 1999 has reached almost 18 billion as on February 2016. At the 2020 horizon, European space missions such as GAIA and EUCLID together with the Large Synoptic Survey Telescope (LSST) will provide a several fold increase of the number of catalogued optical objects while providing measurements of exquisite quality.

Despite the paramount scientific importance of large-scale multi-wavelength observational studies, obstacles still remain for their efficient use by a large majority of astronomers: for example, essential quality information on some catalogues is difficult to gather, and a standard method for a dependable and robust association of a physical source to instances of it in different catalogues (cross-identification) and in diverse spectral ranges is still absent.

Among this plethora of data, the high-energy window to the Universe has strongly benefited from the recent availability of the European Space Agency (ESA) XMM-Newton space observatory. This large X-ray telescope routinely observes the sky with an image quality approaching that of ground-based optical telescopes, generating one of the key ESA astrophysical databases, the third XMM-Newton X-ray catalogue (3XMM), now extending to half a million detections of cosmic X-ray sources, the largest X-ray catalogue ever made.

ARCHES is primarily aimed at increasing the usability and research potential of the XMM-Newton X-ray catalogue, enhancing it with key multi-wavelength resources in a reliable way and distributing these results to the international community, along with the tools that made this possible.

The ARCHES project started by building an enhanced 3XMM catalogue, cleaned of potentially problematic sources, and containing additional useful parameters as well as a

thorough quality assessment. In parallel, ARCHES developed new tools for probabilistic multi-catalogue cross-correlation with extensive archival resources. These so far unique tools allowed us to create high reliability source cross-identifications, based on carefully selected archival astronomical catalogues, so as to cover the widest range in wavelength and sky coverage. These cross-identifications were then used to measure the radiation emitted by each source in the wavelength ranges of all catalogues, building fully calibrated spectral energy distributions (SEDs) of X-ray sources. Another tool specially adapted to extended X-ray sources searches for clusters of galaxies simultaneously in X-ray and optical-near infrared data, with the goal of producing a science-grade catalogue of clusters of galaxies. Most importantly, we validated the quality of these products by using them in the framework of a small number of pilot projects tackling forefront astrophysical topics, providing the route for further enhancement and bringing highly valuable feedback on tools and archival data.

ARCHES tools were offered at the end of the project through dedicated interfaces, while validated multi-wavelength SEDs and associated catalogues were made available to the international community using Virtual Observatory protocols. In this two-fold release of data and tools lies the legacy potential of ARCHES. Our project will pave the way for a full-scale exploitation of the immense wealth of data now (and even more in the near future) available to the international community.

1.3 Description of the main S&T results/foregrounds

1.3.1 Selection of archival catalogues for building spectral energy distributions

One of the first steps of the ARCHES project has been to select among the large collection of catalogues available at the CDS those with the best photometric and astrometric quality while at the same time providing the largest sky coverage. Extracting the final list of catalogues from the several hundred available at CDS has been a major task involving the entire consortium. For each catalogue, we identified the most relevant information, such as the number of sources, the wavelength and the sky coverage. We also obtained a first estimate of the number of expected 3XMM associations and of the source mean positional error. CDS tools such as Aladin, the CDS two catalogue cross-match service, and the multi-order coverage maps (MOC) intersection service were extensively used for that purpose. We later used this list of catalogues, containing several hundred catalogues per energy band (optical, infrared...), and the information gathered for them, to identify in a homogeneous way those catalogues that will be used for cross-matching with the enhanced 3XMM catalogue. The list of archival catalogues was thoroughly revised within the entire consortium and prioritized depending upon the region of the sky considered (mainly high or low Galactic latitude). A special set of catalogues designed for being used by the Integrated Cluster Finder was compiled with the aim to optimally derive photometric redshifts over a range of redshifts as large as possible.

Finally, it became soon clear that the simultaneous multi-catalogue probabilistic cross-matching tool would not be easily usable if the number of catalogues were greater than about seven, due to the geometrically increasing number of combinations. Consequently, we had to severely trim the list of catalogues to be used to 24, defining subsets of them, so that a maximum of 7 (including our "native" enhanced 3XMM catalogue) is used per individual run.

1.3.2 The enhanced XMM-Newton catalogue

The prime objective of building an enhanced 3XMM catalogue (3XMMe) of X-ray sources was to assemble a resource that would be significantly cleaner than the base public 3XMM catalogue from which it was derived and augmented with additional information so as to enhance its value to the ARCHES demonstration science programmes and to the wider user community. The enhanced 3XMM catalogue (designated 3XMMe) is one of the core elements of the ARCHES project, representing the X-ray source basis for the cross-correlations with other multi-wavelength catalogues. The finally released 3XMMe catalogue is a derivative of the latest increment of the 3XMM catalogue, i.e. 3XMM-DR5, that was publicly released in April 2015 tailored for the purposes of the ARCHES project. The 3XMM-DR5 catalogue contains 565962 detections arising from 396910 unique sources drawn from

7781 XMM observations. These numbers include detections from 356 sub-pointings made in mosaic mode.

The construction of the enhanced 3XMM catalogue involves 3 main elements; i) the cleaning (ejection) of detections that are considered to be of lower scientific reliability or quality, ii) the application of criteria to identify detections that are considered potential candidates for the 3 main science themes (i.e. active galactic nuclei (AGN), clusters of galaxies and the galactic plane) of the ARCHES project. The constituent detections of unique sources on the sky are examined to decide which unique sources, overall, meet the criteria to be assigned to each science theme, and iii) the addition of other information that augments the scientific value of the catalogue.

All detections located in bad mosaic fields affected by issues with the ODF (raw) data, those present in flagged high background fields as well as those located at more than 12 arcmin from optical centre were removed. A number of false detections (hotspots) concentrating in localised features when mapped to the separate instrument detector coordinate systems were also discarded. It was also decided i) to exclude fields where the EPIC camera pn detector was off and both MOS1 and MOS2 detectors were off or being used in window modes ii) fields with strong diffuse emission iii) fields with less than 5 ks exposure, iv) pointings for which bore sight correction could not be achieved.

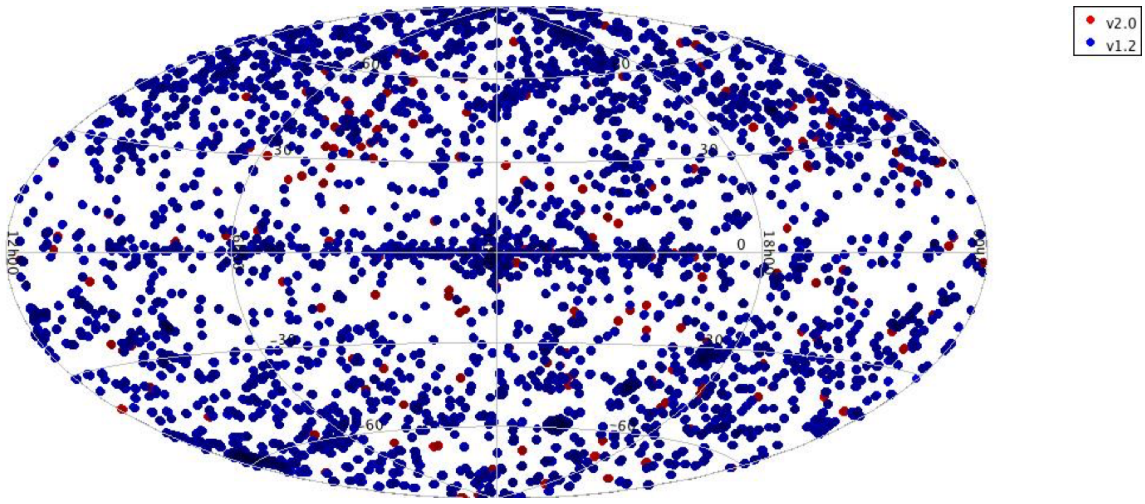


Figure 1: Position of the 4082 XMM-Newton observations entering the 3XMMe catalogue as derived from the 3XMM DR5 (v2.0). We use here the Aitoff projection in which the horizontal axis represents the plane of the Milky Way. The centre of the Galaxy is at the centre of the figure.

Combining these elements, we find that a total of 280757 unique detections are excluded from the enhanced 3XMM catalogue compared to the base 3XMM-DR5 catalogue, leaving 285205 usable detections. These remaining detections are associated with 219788 unique sources and are drawn from 4802 XMM-Newton observations (see Fig.1). It should be

emphasized that although almost half the detections in 3XMM-DR5 are removed from 3XMMe as a result of this cleaning process, most of the rejected detections are perfectly valid detections. The purpose of this cleaning has been to maximise the quality and accuracy of the detections (and unique sources) for the ARCHES project.

1.3.3 A simultaneous multi-catalogue probabilistic cross-matching tool

We have been developing a generic and flexible tool which performs potentially complex multi-catalogue cross-matches and which computes probabilities of association based on a novel statistical framework.

It should be stressed that this development was a theoretical challenge since only sparse literature restricted to the two catalogues case existed at the start of the project. Apart from being able to process an undetermined number of catalogues, the tool had to be flexible enough so as to handle complex list selections on entry, cope with the heterogeneity of astronomical catalogues and be able to deal with various cross-matching algorithms. Another issue was computing efficiency that required the implementation of specific indexation trees.

The complexity of the task and of the hardly predictable progress rate inherent to R&D activities led us to adopt an iterative development scheme and to provide incremental releases of the tool. Test and validation, both from artificial catalogues and from actual cross-matched multi-wavelength catalogues used in the framework of the science cases, provided instrumental and valuable feedback to the software and theoretical development of the tool.

The first task has been to derive a criterion, drawn from statistical considerations, to be used to select a list of potential counterparts to a given source within a set of catalogues. In addition to the criterion applied on the full set of catalogues at one time, we have demonstrated mathematically that the selection can be performed iteratively, i.e. one catalogue at a time, and does not depend on the input order: it is both commutative and associative.

The criterion is based on the positional uncertainties of each individual source. Hence, the search radii on the sky depend on each candidate of each catalogue. To be able to perform such queries efficiently, we conducted a review of the literature in the field of computer science. We have studied, implemented and compared several data structures before choosing the one which best fitted our requirements.

The current version of the software contains more than 25,000 lines of code, implementing several algorithms to compute probabilities up to an arbitrary number of catalogues at the same time.

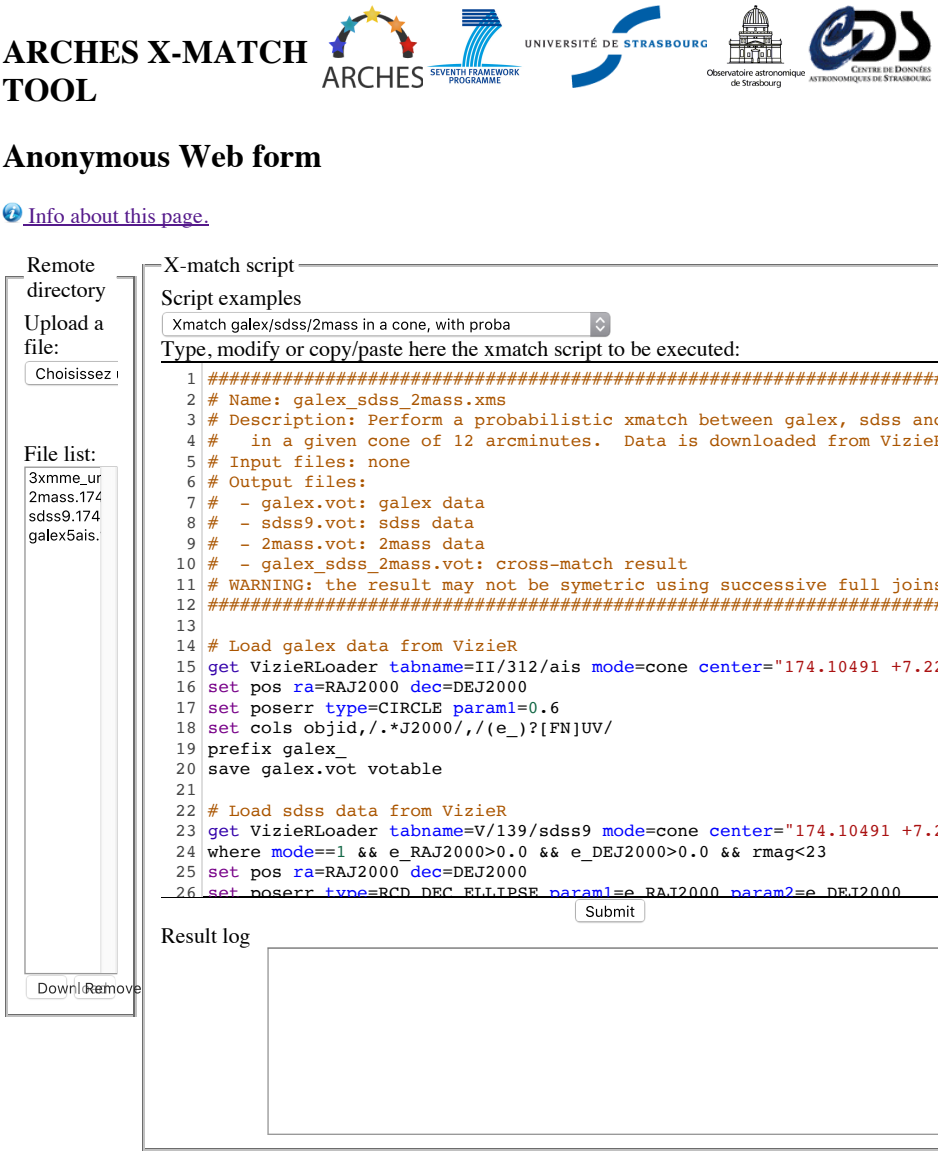


Figure 2: The Web interface to the ARCHES cross-match tool.

The tool runs as a web service on a machine hosted in Strasbourg. On a technological level, it makes use of Java EE with a Tomcat Servlet-Engine running behind an Apache HTTP server. The software runs both from a command line interface and from a web interface (see Fig. 2) implementing a more user-friendly input of the parameters. Its essential functionalities are; i) to load a data table from a file, from the VizieR CDS server, from a service using a VO protocol or available through a given URL, ii) to manipulate the loaded tables e.g. select columns, create new synthetic columns, add metadata, remove rows: one can select the subset of data relevant for their purpose and reject some sources applying potentially complex constraints based on scientific considerations; iii) select the data scientifically relevant for the cross-match to be performed in a way flexible enough to change units on-

the-fly or add systematics, iv) choose one of the 14 cross-match algorithms plus one of the 10 ways to join two tables and v) set the way the relevant information is merged to be ready for the next cross correlation.

An additional software was designed to group the various XMM-Newton fields of view into astrophysically homogeneous regions so as to be able to derive sensible probability priors.

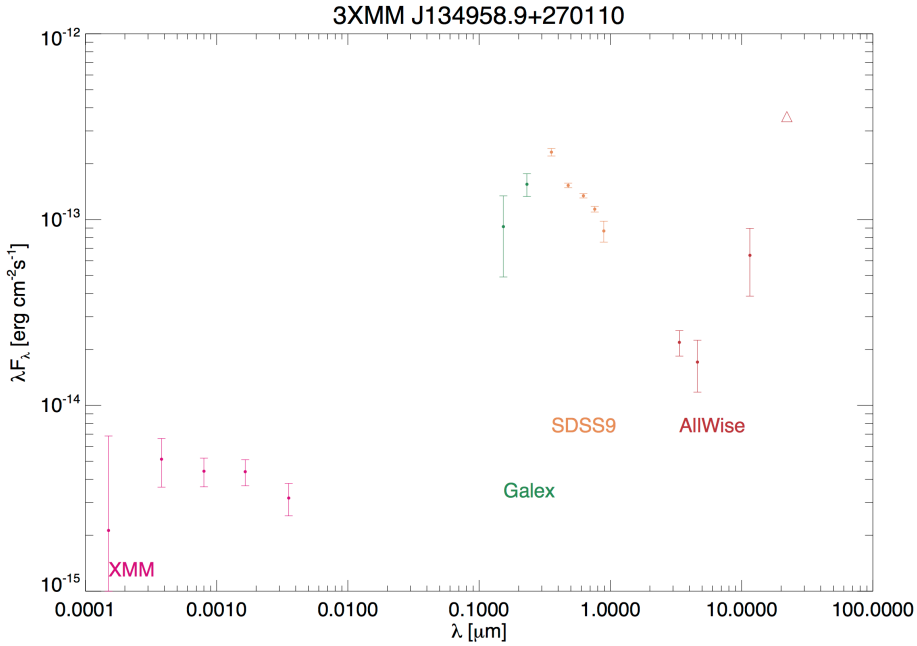
The tool covers a large range of possibilities and allows to test various scenarios beyond those strictly required for the project but which were important for testing purposes and were felt as a valuable and cheap extension of the capabilities of the tool that may further enlarge its usability. For ARCHES purposes we are essentially using both the chi square and the simple cone search algorithms (neglecting proper motions) making the LEFT JOIN of all catalogues starting with XMM.

Although it is theoretically feasible, we raised a point about the practical computation of probabilities when cross-correlating more than 6 or 7 catalogues: a combinatory analysis demonstrates that the number of hypothesis to be considered for a single source increases dramatically with the number of catalogues.

We have developed an original work based on Bayesian probabilities: for the first time the effect of the selection criteria has been introduced in the computation of probabilities for two catalogues cross-matches. We have then resorted to Monte-Carlo simulations to test the equations derived for the two catalogues case and to extend them to three catalogues. The candidate selection of the simulated catalogues has been performed using the first version of the released tool. Hence we have cross-validated both the later one and the new theoretical developments. We have also looked for mathematical simplifications to cope with the computationally extensive problem of prior estimation. The solution we have found reduces an O(NM) problem into an O(N)+O(M).

To our knowledge, it is the first tool allowing to cross-correlate, in a single pass, more than two catalogues containing individual elliptical errors with a coherent statistically based selection criterion.

Together with the Integrated Cluster Finder, The ARCHES multi-catalogue cross-matching tool is at the heart of the ARCHES legacy. It provided multi-wavelength spectral energy distributions for more than 200,000 X-ray sources extracted from the 3XMM DR5 catalogue. Two main catalogues have been produced, one all-sky and another version restricted to the Galactic Plane. All-sky SEDs are based on GALEX 5, UCAC4, SDSS DR9, 2MASS and the AllWISE while the GLIMPSE catalogue is used at low Galactic latitudes. In addition, candidates from the radio SUMSS, NVSS and FIRST and far infrared AKARI FIS catalogues were selected based on a chi-square criterion.



3XMMe Coordinates:	13 49 58.9 +27 01 10 ± 0.92arcsec
Merged Coordinates:	13 49 58.9 +27 01 10 ± 0.10arcsec
3XMM Identifier:	SRC_ID=202051902010024
Detection Max Likelihood:	162.05
Variability Flag:	true
Arches Catalogue:	xmatch_noproba_withflux_xgustwra_v2.01.vot
Arches Identifier:	uid=209

Most likely association pattern (coloured points - upper limit values are marked by triangles):
[XMM, Galex, SDSS9, AllWise] proba(1) = 0.98

No other patterns with at least one association (proba threshold = 0.01)

Source coverage: [XMM, 2MASS, Galex, Ucac4, SDSS9, AllWise, Nvss, FIRST+NVSS]

Figure 3: A graphical representation of the spectral energy distribution of an X-ray source extracted from the 3XMMe catalogue. Among other information, the page lists the names of the archival catalogues from which the various fluxes are extracted as well as the overall probability of the total combination (here 98%).

1.3.4 Creating spectral energy distributions

One of the main ARCHES objectives was to produce a scientifically validated, VO-compliant set of spectral energy distributions of the objects included in the enhanced 3XMM-Newton Serendipitous Source Catalogue (see Fig. 3). Transforming magnitudes and errors into absolute fluxes, errors and determining central wavelengths, all was performed using the Spanish Virtual Observatory Filter Profile Service. In addition, a database was created in order to test and validate the SEDs and provide a world access to these data by implementing VO protocols.

1.3.5 An integrated cluster finder

One of the ARCHES's objectives was to develop a software tool, dubbed the Integrated Cluster Finder (ICF), able to search for galaxy clusters at the position of the extended X-ray sources registered in the XMM-Newton catalogue (and virtually in any direction of the sky) using ground- and space-based multi-wavelength imaging and spectroscopic data.

Extended X-ray sources located at high galactic latitudes are the tracers of clusters of galaxies. X-rays are emitted by the hot intergalactic matter trapped in the deep potential well provided by the dark matter mass assembly. While clusters of galaxies are single entities at X-ray wavelengths they appear as a group of objects at optical/infrared wavelength. There, only the individual member galaxies with largely identical redshift can be recognized. While the galaxies do not play a major role in the mass budget of a cluster, they are of utmost importance for assessing the reality of the cluster through a measured over-density of galaxies and for redshift determination. The first aspect is important to safeguard against spurious extended sources, based e.g. on imperfections of the telescope optics or of the detection process. The cosmological redshift is imprinted into spectral features of the member galaxies and/or the colour of the member galaxies and those properties are the basis of the detection process with the ICF.

The ICF was used to make the important step from a catalogue of detections in multi-wavelength data to a science-grade catalogue of galaxy clusters.

We restrict ourselves to high Galactic latitudes in order to avoid problems with source confusion in the Galactic plane and to avoid problems due to source extinction in the Milky Way.

The ICF implements a redMaPPer-like algorithm to search for clusters in the space spanned by the positions and the colours of their member galaxies. A major difference with respect to the published redMaPPer method is that prior information on the position of the cluster from the detected X-ray position is integrated in our method.

The cluster finding approach is based on the fact that luminous red galaxies (LRGs) have (on the average) different colours at different redshifts. From a calibrated redshift-colour relation one can thus estimate the redshift of a member galaxy and, via summation over all potential member galaxies, the redshift of the cluster. The algorithm contains three main steps: i) calibration of the redshift-colour relation: for a chosen set of external catalogues (e.g. SDSS, UKIDSS, AllWISE) one obtains average colours and colour-colour covariance matrices for each redshift. An alternative version of the ICF that uses pre-fabricated values of photometric redshifts (like those provided with SDSS, CFHTLS or ALHAMBRA catalogues) has also been developed, ii) estimation of the background density to be incorporated into the determination of a multiplicity (abundance of objects) at the cluster position and iii)

cluster search proper: Identification of possible members and estimation of the redshift of each cluster (see Figs. 4 & 5).

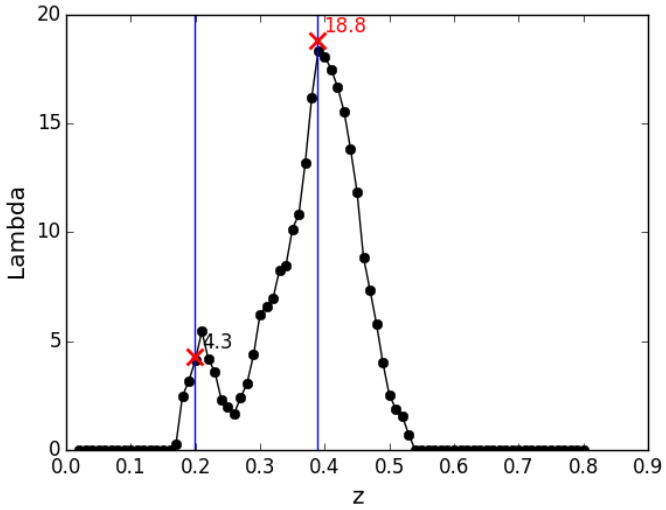


Figure 4: Measure of the probability of finding a genuine cluster versus redshift

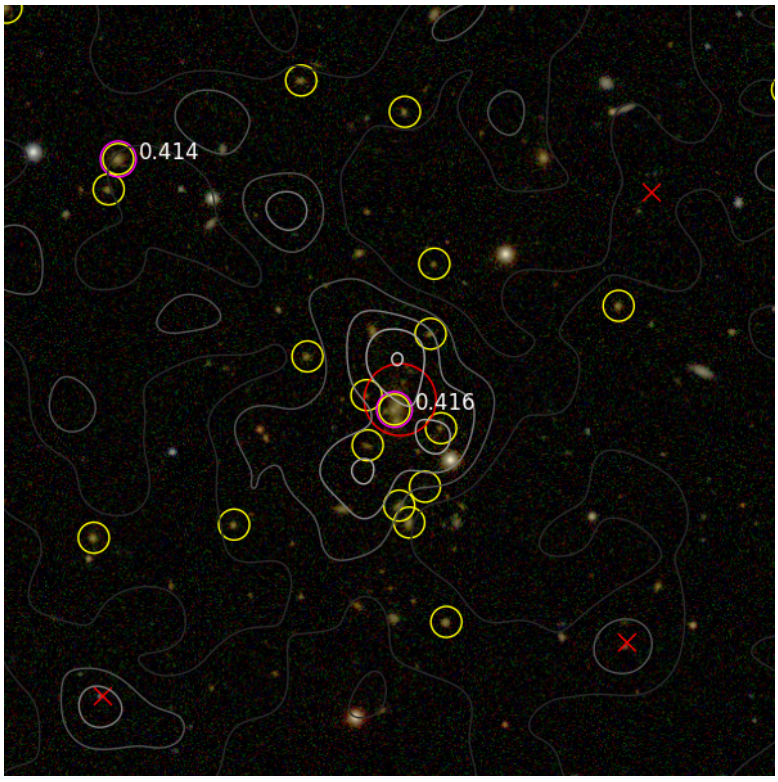


Figure 5: Corresponding finding chart. Coloured 2' by 2' SDSS image overlaid with XMM X-ray contours. Red circle: X-ray source position and extent. Red crosses: point-like X-ray sources. Yellow and light blue circles: likely cluster members with spectroscopic redshift indicated by purple circle and label where possible.

Data from various input catalogues in a circular field of radius R = 8 arcmin are acquired around the position of each extended 3XMMe source. Therefore, our approach is inefficient for the detection of clusters at low redshift as those appear larger than 8 arcmin on the sky. This radius corresponds to 1 Mpc co-moving radius at redshift z = 0.12. We note, however that this limitation does not question the validity of our search, as most nearby clusters (z < 0.12) are already known. Also, bright, nearby (hence very extended) clusters are not reliably detected as single entities by the X-ray source detection chain and are thus partially screened away by the filters applied to generate 3XMMe.

The ICF has been tested successfully against several samples of identified clusters of galaxies. Applied to the enhanced 3XMM catalogue produced by ARCHES, the ICF yielded a catalogue of cluster candidates with 481 detections.

1.3.6 Test science cases

A set of seven distinct pilot science cases were selected at the beginning of the project in order to investigate the usability of ARCHES tools and products in a range of science themes from planetary formation to cosmology. ARCHES designed a solid testing and validation programme which guaranteed the quality and robustness of the tools. However, as anticipated in the project, using ARCHES products in the framework of science cases provided strong feedback which allowed us to add new features and revealed subtle issues that would have naturally escaped the scrutiny of the main stream testing activities. For instance, AGN studies uncovered a notable problem affecting astrometry in the 3XMM-DR4 catalogue. This problem was conveyed to the Survey Science Centre, the consortium responsible for the production of the XMM-Newton catalogues.

1.3.6.1 Active Galactic Nuclei (AGN)

The AGN science cases encompass two paths of investigations, a characterization of the multi-wavelength properties of obscured AGN and a study of the relation between jet and overall AGN properties in radio-loud objects.

i) Obscured AGN

We have concentrated on using ARCHES tools to trace the obscured AGN phenomenon in the optical, X-ray and mid-infrared (MIR) regimes and to compare the obscured AGN populations identified at different wavelengths.

We have used sources in a subset of 3XMMe fields away from the Galactic plane and completely covered by the Sloan Digital Sky Survey (SDSS). In these fields, X-ray (X), SDSS (S, optical) and WISE (W, infrared) were cross-correlated, building source samples having different combination of association probabilities between catalogues (see Fig. 6). The use of

the ARCHES cross-correlation tool has been instrumental in this endeavour, since it allows matching three catalogues providing full probability information in all combinations. We have then used only those SDSS sources with significant and luminous [OIII]5007Å emission, to select AGN (although a contribution of strong star formation -SF- could also contribute substantially).

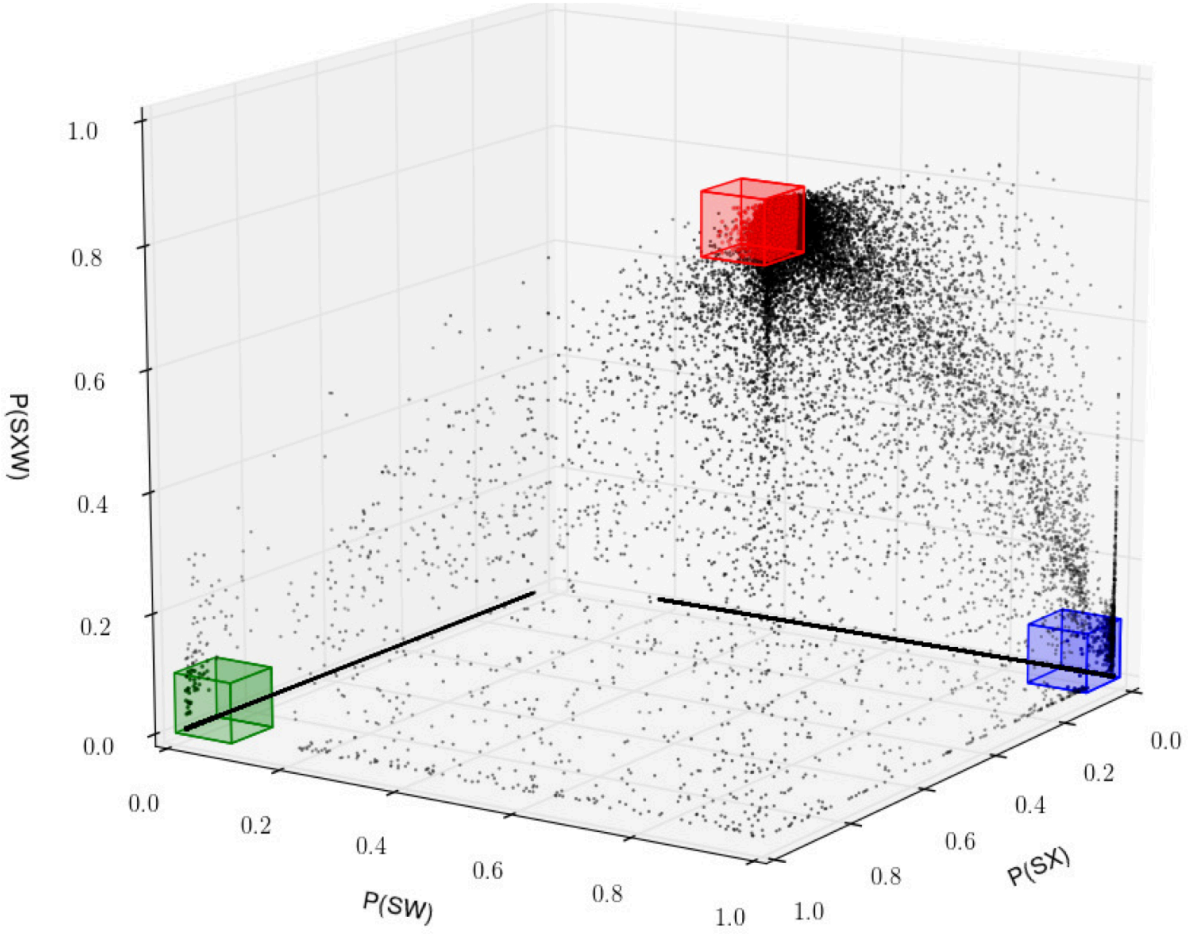


Figure 6 : Selection of sources in X-ray (X), optical (S) and infrared (W): the axes are probabilities of association so that the red volume indicates the sources with high probability (>90%) of being detected in the three bands; the green area shows the sources with high probability of SX association, but no W detection (<10%); the blue area highlights the sources with high probability of SW association, but no X detection. This allows investigating what types of AGN are preferentially detected in each band, to gain insight into their intrinsic properties

For the X-ray sources with SDSS counterparts but no infrared (W) match, we obtained MIR fluxes using the published forced photometry on the SDSS positions and for the sources matching both the SDSS and WISE but void of any X-ray detection, we obtained X-ray upper limits from the FLIX server. We finally cross-correlated the three sub-samples with 2MASS and UKIDSS to obtain near infrared photometry. We have fitted the SEDs and developed a

method to obtain X-ray spectral information from the 5-band count rates in the 3XMMe catalogue. The result is a large sample of > 1500 objects with wide and uniform optical-to-MIR SED and X-ray spectral information. Our first result is that there are almost no AGN that are undetected in MIR/WISE at the typical depths of 3XMMe.

In the first place, we use the Baldwin, Phillips & Terlevich diagram that is essentially a 2D plot of ratios of two pairs of narrow emission lines. Most sources detected in X, S and W are consistent with AGN while those not detected in X-rays are likely star forming galaxies. However, some of the non X-ray detected sources are in the AGN region. The X-ray spectral information shows that this could be because they are about 0.5dex less X-ray luminous for their MIR luminosity or because they are absorbed by typical column densities of ~10^{24}cm^{-2}.

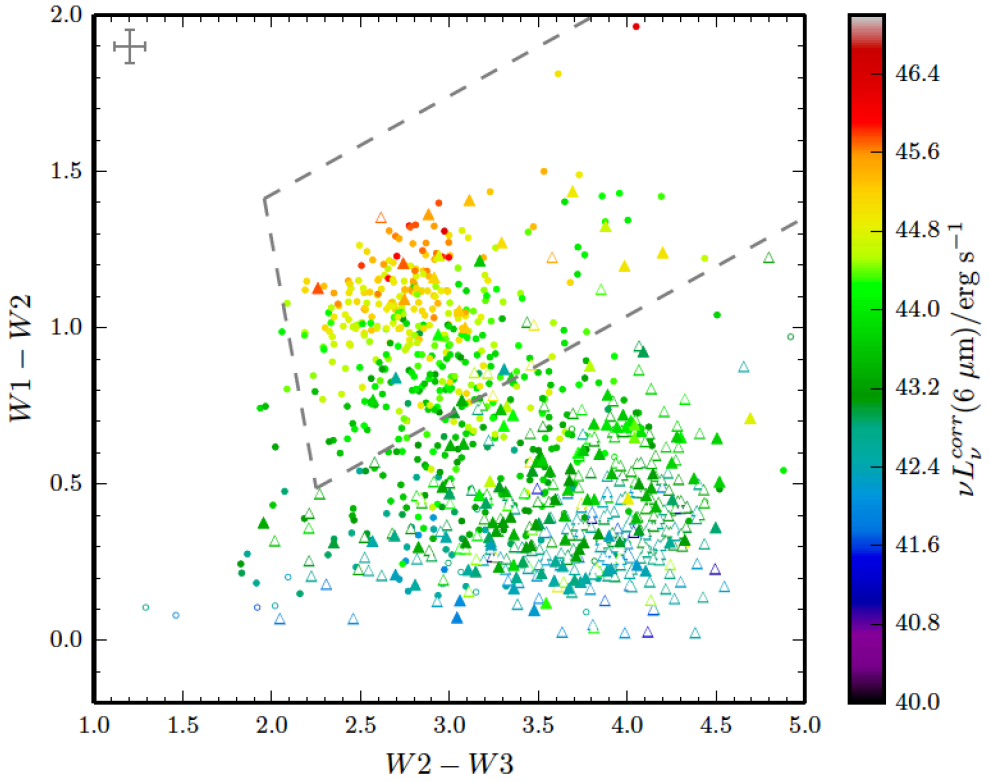


Figure 7: Selection of AGN using MIR-colours, showing essentially the logarithms of the ratios between WISE bands (2nd and 3rd band along the X-axis and 1st and 2nd along the Y-axis). The dashed grey line delimits the area where AGN are expected to lie (Mateos et al. 2012, this project). The colour of the symbols correspond to the MIR luminosity and show clearly that this criterion is effective in selecting luminous AGN, but fails to select less luminous ones, probably because the contribution from their host galaxies change their MIR-colours

We have also examined an alternative method to select AGN, which uses MIR colours and is based on the effective power-law shape of the SED of AGN in that band. Again as expected, most sources detected in S, X and W are in the AGN region and most of those not detected

in X-ray are outside it. Our multiple criteria to define AGN have also allowed us to estimate how efficient is the MIR AGN selection method as a function of the MIR luminosity.

Finally, we have compared the numbers of sources classified using each of the two criteria above (plus a third one related to the intrinsic X-ray luminosity: if it is above 10^{42} erg s^{-1} in the 2-10keV band, the source must be an AGN, because those luminosities are unattainable with star forming alone).

We have concluded that, for serendipitous XMM-Newton surveys, the combination of catalogues and wavelengths that is best suited for AGN studies is SDSS in the optical, WISE in the MIR and UKIDSS in the Near-IR, combined with 2MASS in the latter band for the brightest sources and/or for wider coverage. We have also used a variety of SED-fitting methods but, in the end, we have decided to develop our own Bayesian method in Python, using the CIAO-Sherpa package and some Markov Chain Monte Carlo libraries, which combined a minimum of indigenous code writing with flexibility in the choosing of the template libraries.

ii) Radio-loud AGN

We have also compiled and exploited a sample of radio-loud AGN, in particular, demonstrating the value of radio emission to identify the presence of AGN where the host galaxy emission dominates in the mid Infrared.

Using the ARCHES cross-matching tool, we built the MIXR (Mid-Infrared, X-ray, and Radio) sample by combining the WISE, 3XMMe and FIRST+NVSS source catalogues. MIXR is the largest, uniform sample to date at these wavelengths, covering a wide range of galaxy properties, from starburst to quasars. With our selection we can identify AGN activity even in sources where the host emission dominates, and efficiently diagnose star formation, thus allowing us to derive reliable diagnostic classifications that can be used in other samples. The study sought to investigate the relationship between radiative and jet output in radio-loud AGN, and its implications for regulating mechanisms for the jet and long-term variability timescales. The results show potential problems in our current understanding of the relationship between AGN and star formation activity: if both the radiative and the jet output of an AGN can vary by 4-5 orders of magnitude on timescales of a few million years, and the AGN influence on the star formation in its host also happens in similar timescales, it may not be possible to study the relation between them before we better understand the life cycles of AGN.

The main objectives of our study involved: i) using the MIXR sample to devise a series of effective diagnostics that will allow the community to pre-classify extragalactic sources even in the absence of redshift, based on their radio, mid-IR and X-ray fluxes. This will establish a baseline that might facilitate source triage when the next generation of instruments (JWST,

E-ELT, Athena, ALMA, SKA...) start producing vast amounts of data, ii) using redshifts from SDSS DR12 to verify the accuracy of the early diagnostics, and explore AGN activity versus star formation contribution in sources dominated by their hosts, examine the apparent lack of correlation between kinetic (jet) and radiative output in radio-loud AGN, to draw constraints on its possible origin and its implications on the accretion properties of AGN, and study the selection biases that arise from the selection methods in our sample, and in others, particularly AGN samples with a mid-IR/X-ray selection.

1.3.6.2 Galactic Science cases

We have worked in four independent projects in order to learn about the nature of the X-ray Galactic landscape.

i) X-ray emission mechanisms in the high mass star population of the Galaxy, wind colliding binaries and Gamma-Cas analogues

The nature of the low- to intermediate-luminosity ($Lx \sim 10^{32-34}$ erg s^{-1}) source population revealed in hard band (2-10 keV) X-ray surveys of the Galactic plane is poorly understood. To investigate their nature, we cross-correlated the XMM-Newton 3XMM-DR4 survey with the infrared Two Micron All Sky Survey and Galactic Legacy Infrared Mid-Plane Survey Extraordinaire (GLIMPSE) catalogues. We identified reliable X-ray-infrared associations for 690 sources. We selected 173 sources having hard X-ray spectra, typical of hard X-ray high-mass stars (kT > 5 keV). Roughly 15 per cent of the hard sources are classified in the literature: ~68 per cent as high-mass X-ray stars single or in binary systems (WR, Be and high-mass X-ray binaries - HMXBs), with a small fraction of G and B stars. An infrared spectroscopic pilot observation at the William Herschel Telescope allowed us to identify five hard X-ray sources with Wolf-Rayet stars, supergiant stars, Be Gamma-Cas analogues or possible interlopers. Taking advantage of the success of this pilot survey we started in 2016 an extended observing campaign covering the central part of the Galactic disc.

ii) A study of the X-ray emitting Wolf-Rayet stars

The space density of Wolf Rayet (WR) stars is still not well constrained. Nor is the mechanism producing copious X-ray emission in some of these objects. While the expected number of single and binary WRs in the Galaxy varies from 1200 to 6500, only about 640 Galactic WRs are known up to date. Moreover, the origin of the sometimes-observed hard X-ray emission in WRs is highly debated. There are three competing scenarios: 1) intrinsic emission, where instabilities in the radiative wind could produce hot shocks 2) accretion onto a compact object, or 3) wind-wind interaction in a binary star (WR+O, WR+WR).

In order to learn about the X-ray emission mechanism and to constrain the space density of WRs we need a large and unbiased sample. For that purpose, we cross-correlated the most recent catalogue of WRs in the Galaxy with the ARCHES catalogue. To ensure that we have

the right X-ray counterpart for each WR we visually inspected the ARCHES SEDs and the finding charts for all the WRs having an X-ray candidate. We investigated the fraction of WR stars as a function of spectral type and whether this varies with X-ray activity. We also investigated a possible relation between X-ray flux and orbital period for known WR binaries.

Iii) Modelling of the solar type stellar populations in X-rays.

In stars such as the Sun, X-ray emission is closely related to the existence of convecting cells whose activity depends on differential rotation. Since stars spin down due to magnetic braking, X-ray emission is a good tracer of the age of the stars and reveals the younger ones. The aim of this study is to derive the main population properties of young stars, their space density (to derive the local star formation rate), scale height (to constrain the mechanism that scatters young stars away from the Galactic disc) and to derive the importance of the old population whose X-ray activity is maintain by binary rotation.

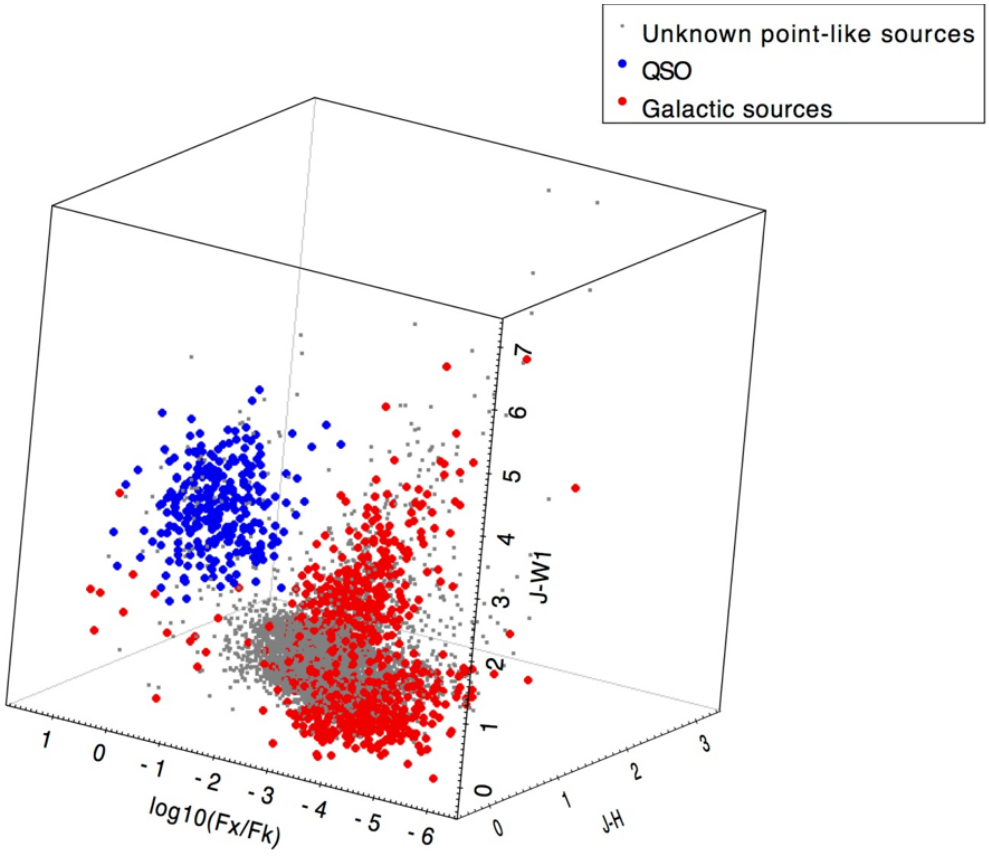


Figure 8: Groups of identified active galactic nuclei (AGN, blue dots) & galactic stars (red dots) separate well in the Fx/Fir + IR colours space. Sources of unknown nature (small grey dots) Unknown can easily be classified in AGN or galactic stars using for instance a kernel density method.

We selected sources in the ARCHES catalogue having optical and infrared counterparts. A learning sample based on published catalogues containing stars with known properties was compiled. Making use of this learning sample, and through a principal component analysis, we classified sources as either stars, quasars or galaxies (see Fig. 8). For stars, by fitting stellar atmosphere models to the observed SED, we determined the stellar parameters, such as effective temperature and gravity. We validated the obtained results by comparing them with spectroscopically determined values taken from the literature. The modelling of the stellar X-ray population will be done using the X-ray count model developed by Guillout et al. 1996 and based on the evolution synthesis population model from Besancon. This will provide us with the surface density of sources detected as a function of X-ray flux, age and spectral type, at any direction of the Galaxy as a function of scale height, stellar formation rate, etc... to be compared with our observations.

iv) Circumstellar discs around late-type stars

Young late type (F-M) stars are known to be copious X-ray emitters. Their X-ray emission arises from the stellar corona, and it is directly related to the rotation period of the star. The fast rotation is usually the effect of the heritage of the rotation of the parental cloud from which the star was created, due to angular momentum conservation.

While gas discs are found around young stars as they dissipate after 10-20 Myr, debris discs' remain with lifetimes spanning over Gyr (e.g. the Kuiper Belt in the Solar System). Although the emission of debris discs depends on several parameters, it is expected that the fractional emission of a debris disc decreases with stellar age. In this work, we use the X-ray luminosity (Lx/Lbol) to calculate the stellar age and we search for (expected) relations between the stellar discs and the stellar age.

A total of 108,037 sources not affected by bright star or galaxy flag were selected from the ARCHES catalogue and we cross-correlated them with UCAC4 to obtain proper motion information. After imposing good photometric quality and a star flag in the UCAC4 catalogue we obtained > 10,000 objects. We then selected only those with significant proper motions (pm > 3σpm), obtaining a sample of ~ 5800 galactic candidate sources.

We used the Virtual Observatory tool VOSA to identify infrared excess in ~600 candidates. In order to discard any source known to be not a single star, we cross-matched our sample with the SIMBAD database. 154 objects were removed from the sample.

Stellar and disc parameters of the remaining ~ 400 sources were derived. These parameters are presently being analysed before obtained the final list of bona-fide young, late-type stars with debris disc.

1.3.6.3 Clusters of galaxies

The Integrated Cluster Finder (ICF) was used to generate a catalogue of clusters based on the enhanced 3XMM catalogue. The very first important test to be performed was the robustness of the ICF redshift determination with respect to reference cluster samples either collected in this project of found in the literature. The ICF passed this test with a confidence between 95% and 98% (see Fig. 9). Clusters up to redshift one were reliably recovered from published samples.

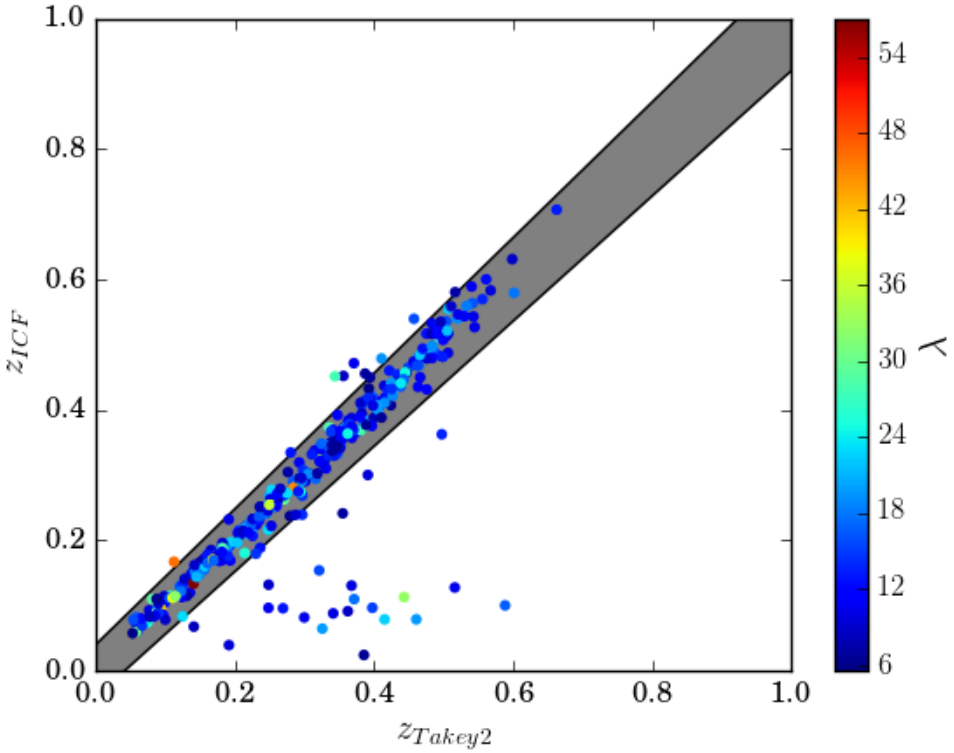


Figure 9: Comparing the redshifts of 3XMM clusters candidates identified by the ARCHES Integrated Cluster Finder with those derived in former works, here the catalogue of Takey et al. 2013.

Cluster catalogues based on the SZ-effect from the Planck satellite or from the South Pole Telescope became available during the course of the project. These catalogues contain predominantly massive clusters, given the sensitivity of the current SZ-instruments. Newly discovered XMM-Newton clusters on the relatively small survey area however are mainly low mass objects. The overlap between current SZ-cluster catalogues and those from XMM-Newton is small, just five clusters were found jointly in both catalogues. The overlap is excessively small to derive meaningful conclusions from a comparison.

The behaviour of the ICF with respect to false positives and to completeness was characterized. A paper describing the method of cluster identification, redshift

determination together with a description of the newly established cluster catalogue was submitted to A&A.

For all detected clusters based on the enhanced version of the 3XMM-DR5 catalogue finding charts were produced. These show X-ray contours on top of an optical image from the SDSS survey with likely member galaxies of the cluster marked as well.

The X-ray images of all new cluster candidates were inspected and the X-ray spectra extracted. X-ray spectral fits were performed for those clusters with reliable redshift information to determine the cluster temperature and the cluster luminosity.

A luminosity-temperature diagram was generated and various strategies were tested to optimally fit the underlying relation taking biases into account. At the same time, a dedicated follow-up program for high-redshift clusters beyond the SDSS limit has started that already allowed us to identify 7 clusters with redshifts between 0.84 and 1.22.

1.3.7 Disseminating ARCHES products

ARCHES success also rests on the quality, user friendliness and efficiency of the interfaces to data. ARCHES data can be accessed in four flavours, as flat files, through specific and rich interfaces designed by the project, through the Virtual Observatory, and in 2016 through standard services at the Centre de Données astronomiques de Strasbourg (CDS).

Both cross-correlated catalogues and the enhanced 3XMM catalogues can be downloaded as compressed VO tables from the ARCHES website.

ARCHES designed a specific interface, the ArchesDB, to browse, expose and select the cross-correlated catalogues, the resulting spectral energy distributions and the catalogue of cluster candidates created by applying the Integrated Cluster Finder to extended X-ray sources in the enhanced 3XMM catalogue (see Fig. 10). The ArchesDB is an extension of the 3XMM interface of the SSC, the XCatDB, augmented with ARCHES data. Consequently, the ArchesDB proposes a rich Web interface enabling the users to run very complex queries and to easily browse data thanks to number of previews of spectra, SEDs, light curves, etc...

The spectral energy distributions (SEDs) are also made available through an archive interface developed by SVO, namely the CAB-INTA database.

The enhanced catalogue and the cross-matched catalogues will also be published in VizieR at CDS in 2016, awaiting the release of the scientific publications. SEDs will also be stored in Vizier in order to include them in the long-term preservation policy of the CDS.

A number of Virtual Observatory interfaces are implemented. The catalogue is published into the VO as simple cone search resources (CSP) by the ArchesDB and soon in Vizier as any other Vizier catalogue. The SSAP VO interface (Simple Spectrum Access Protocol) interface for the SEDs comes with the SVO archive and from Vizier in the near future. The ArchesDB offers a TAP (Table Access Protocol) access to the ARCHES Data with an ObsTap (data model uses for data discovery) implementation. A multi-scaled X-ray image (HIPs) covering the ARCHES sky has been made available through Aladin, the VO sky atlas. It can also be used to help browsing the data and selecting regions of interest on others interfaces.

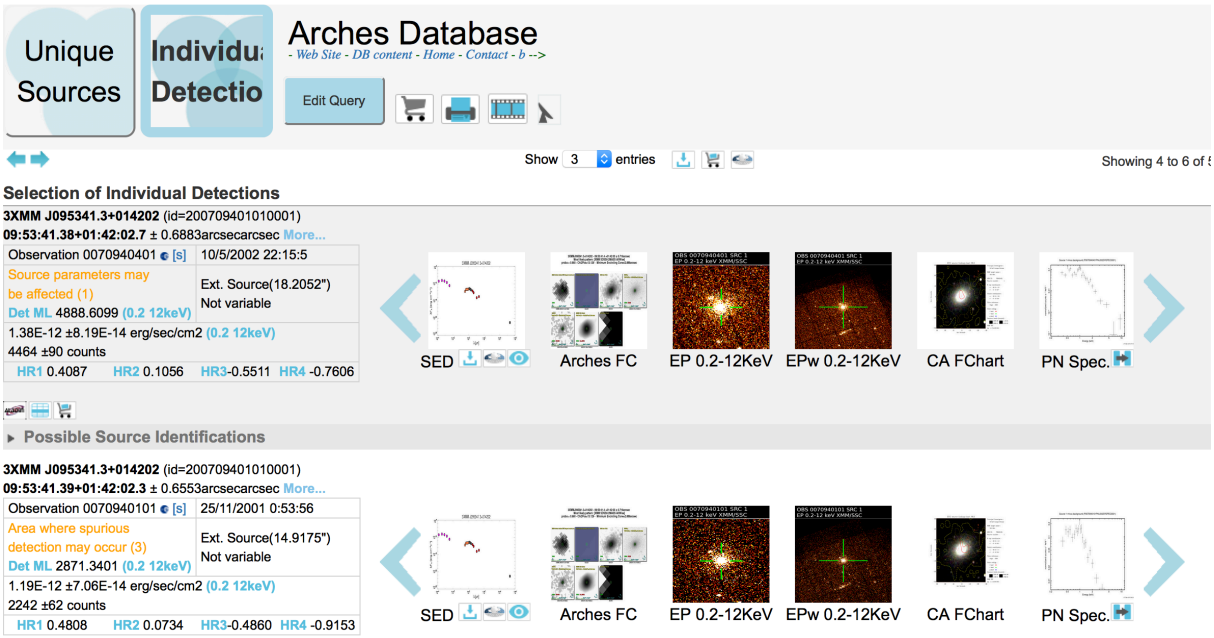


Figure 10: Screenshot of the Arches Database. The interface allows the user to select 3XMMe sources based on simple and complex criterions. For each selected entry, the interface displays thumbnails giving access to the relevant data (spectral energy distributions, finding charts in several bands, X-ray images, X-ray spectra, X-ray time series, etc..). Source parameters and identifications at other wavelengths are also listed.

Outreach:

Apart from outreach pages exposed on the ARCHES web site, the project developed an original and versatile outreach device dubbed as the "Arches Walker". This project aims at helping the public to discover by themselves different aspects of the ARCHES scientific targets. The Arches Walker is built on the top of a data corpus containing the description of a sample of sky objects relevant for the ARCHES science. The user can select and display objects in various wavelength. He/she can also explore their neighbourhood or get some textual data. Two interfaces are available so far: a Web application included within the ARCHES Web site, and a device intended to be used on exhibition booths that use a tablet as a remote controller for the display (see Fig. 11).

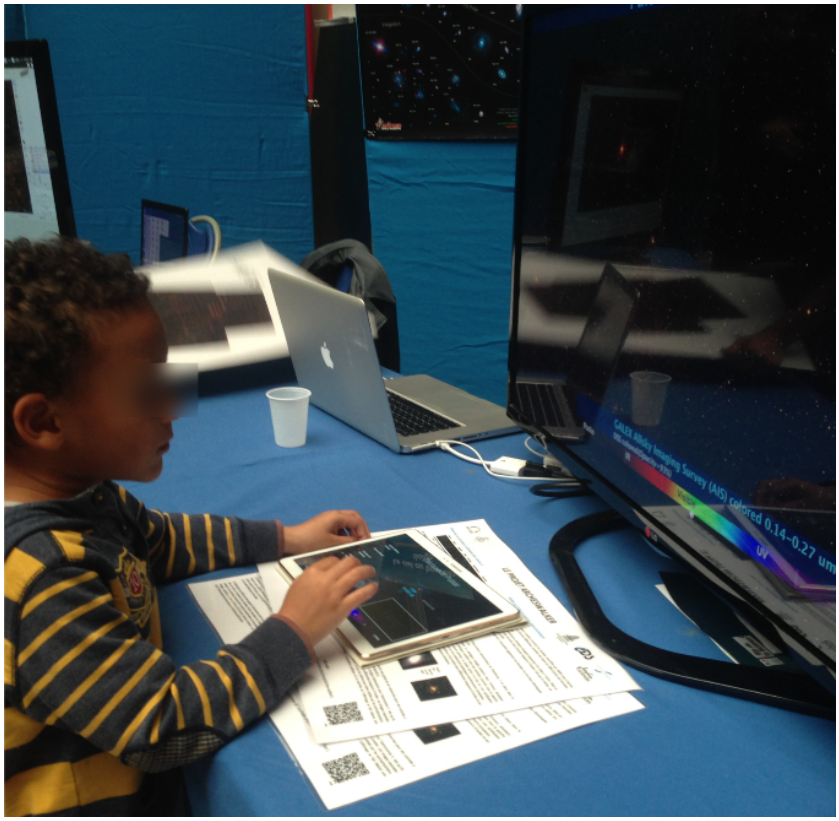


Figure 11: A young future astronomer discovering the beauty of the multi-wavelength sky using the ARCHES walker.

1.4 Potential impact

The objective of ARCHES was to build scientifically-validated spectral energy distributions for the many stars, galaxies and clusters of galaxies detected by the ESA X-ray space observatory XMM-Newton. These multi-wavelength enhanced data sets will greatly facilitate the opening of new avenues of investigation and therefore need to be widely advertised. By bringing together expertise from leading institutes in the field of multi-wavelength analysis, ARCHES has opened to investigation new exciting domains of research and will very significantly increase the overall scientific return of this highly successful European space observatory at a modest cost.

ARCHES data and products are distributed to the international community through a variety of channels that altogether will unfold the impact of ARCHES work. ARCHES cross-correlated catalogues and spectral energy distributions are published in the Virtual Observatory that is becoming a standard data access mode for the astronomical community. In parallel, these data can be accessed through the ARCHES web site, through dedicated interfaces (the ArchesDB and the CAB database) exposing in a handy manner ARCHES data and their relations to other databases. From mid-2016 on ARCHES data will be also accessible through the Vizier catalogue service of the Centre de Données astronomique de Strasbourg (CDS), a world reference for astronomical data curation and distribution.

As far as tools are concerned, the Integrated Cluster Finder and the ARCHES cross-correlation tool are in open access on a server located and maintained in Strasbourg. More importantly, the simultaneous multi-catalogue cross-matching program will be offered by mid 2016 as an option in the successful cross-matching service developed at CDS. This Web service is used worldwide and accessible through the main astronomical data analysis and handling software such as TOPCAT.

The XMM-Newton satellite is the most scientifically productive European X-ray observatory ever flown and is currently one of the flagships of the European Space Agency's astronomical programme. On board consumables allow operations to be pursued until 2025 provided ground segment funding remains close to the current level and provided detectors are kept in good health. Keeping this X-ray satellite in operation as long as technically possible would allow the European community to minimize the gap between XMM-Newton and ATHENA, the next large European X-ray telescope to be launched around 2028.

In parallel, the German-Russian eRosita X-ray telescope scheduled for launch in 2017 will carry out the first wide energy range (0.5-10 kev) all-sky survey, thus providing an order of magnitude more sources than the ROSAT survey operated in the nineties.

ARCHES products already constitute a solid basis for carrying out forefront astrophysical research using data accumulated during the first 15 years of XMM-Newton operations.

However, ARCHES's impact will very certainly extend much beyond the date of products release. From the very beginning, one of the main concerns of the project was to ensure that enough expertise would remain close to the participating institutes so as to maintain the tools in operation and possibly develop them further in order to implement new applications. As stated above, the ARCHES probabilistic cross-matching tool will be soon included in the array of CDS web services under the responsibility of the code developer, recently hired on a fixed position. Although the situation is less clear for the ICF, there are good prospects to keep a reasonable level of maintenance in the forthcoming years.

It is important to note that the ICF is not restricted to XMM-Newton data. In fact, the tool can be used at virtually any location of the sky. It has been therefore envisaged to use the ICF to detect new clusters of galaxies in the former ROSAT all-sky survey and in the eRosita all-sky survey that will be released around 2020.

The same remark applies to the ARCHES cross-matching tool. Being not specific to X-rays, the algorithm offers a wide range of potential applications. Proposed as a worldwide accessible CDS service, it is expected to play a major role for building the spectral energy distributions of the billions of astronomical objects that will be harvested at the horizon 2020 by GAIA, EUCLID, the LSST, SKA, etc.…

List of ARCHES refereed publications

- “Infrared identification of hard X-ray sources in the Galaxy “, A. Nebot Gómez - Morán, C. Motch, F.-X. Pineau, F. J. Carrera, M. W. Pakull, F. Riddick, 2015, MNRAS, 452, 884; DOI: 10.1093/mnras/stv1020
- “The 2XMMi/SDSS Galaxy Cluster Survey. III. Clusters associated with spectroscopically targeted luminous red galaxies in SDSS-DR10”, A.Takey, A. Schwope, G. Lamer, 2014, Astronomy & Astrophysics, Volume 564, id.A54, 19pp, DOI : 10.1051/0004-6361/201322973
- “Uncovering obscured luminous AGN with wise”, S. Mateos, A. Alonso - Herrero, F. J. Carrera, A. Blain, P.Severgnini, A.Caccianiga and A. Ruiz”, 2013, MNRAS, 434, 941; DOI: 10.1093/mnras/stt953
- “The 2XMMi/SDSS Galaxy Cluster Survey. II. The optically confirmed cluster sample and the L_X-T relation”, A.Takey, A.Schwope, G. Lamer, 2013, Astronomy & Astrophysics, Volume 558, id.A75, 17 pp, DOI : 10.1051/0004-6361/201220213
- “The XMM-Newton Bright Survey sample of absorbed quasars: X-ray and accretion properties“, Ballo et al., 2014, MNRAS, 444, 2580; DOI: 10.1093/mnras/stu1628
- “Exploring the active galactic nuclei population with extreme X-ray-to-optical flux ratios (fx/fo > 50)”, Della Ceca et al. 2015, MNRAS, 447, 3227; DOI: 10.1093/mnras/stu2665
- “Submm-bright X-ray-absorbed QSOs at z ˜ 2: insights into the coevolution of AGN and star formation”, Khan-Ali et al. 2015, MNRAS, 448, 75; DOI: 10.1093/mnras/stu2719
- “Revisiting the relationship between 6 µm and 2-10 keV continuum luminosities of AGN”, Mateos et al. 2015, MNRAS, 449, 1422; DOI: 10.1093/mnras/stv299
- “WISE colours and star formation in the host galaxies of radio-loud narrow-line Seyfert 1”, Caccianiga et al. 2015, MNRAS, 451, 1795; DOI: 10.1093/mnras/stv939
- “The Integrated Cluster Finder for the ARCHES project”, Mints et al. 2015, A&A, Submitted.

Other ARCHES related contributions

- “Astronomical Resource Cross-matching for High Energy Studies”, poster by C. Motch at the ADASS XXV conference that took place in Sydney on 25-29 October 2015. ASP in press
- “Radio-loud AGN through the eyes of 3XMM, WISE and FIRST/NVSS”, talk by B. Mingo at the "Demographics and environment of AGN from multi-wavelength surveys" conference that took place in Chiana (Crete) on 21-24 September 2015.
- “The 3XMM-DR4 of Wolf Rayet stars”, talk by A. Nebot at the International Workshop on Wolf Rayet Stars in Potsdam, 1-5 June 2015.
- “The ARCHES project”, poster by Motch et al., Astronomical Data Analysis Software an Systems XXIV (ADASS XXIV), Proceedings of a conference held 5-9 October 2014 at Calgary, Alberta Canada. Edited by A. R. Taylor and E. Rosolowsky. San Francisco: Astronomical Society of the Pacific, 2015., p.437
- “Towards a next-gen catalogue cross-match service”, talk by Pineau F.X.; Boch, T.; Derriere, S. and the Arches Consortium, Astronomical Data Analysis Software an Systems XXIV (ADASS XXIV), Proceedings of a conference held 5-9 October 2014 at Calgary, Alberta Canada. Edited by A. R. Taylor and E. Rosolowsky. San Francisco: Astronomical Society of the Pacific, 2015., p.61
- “Radio-loud AGN through the eyes of 3XMM, WISE and FIRST/NVSS”, poster by B. Mingo at the X-ray Universe 2014 conference, 16-19th June, Dublin.
- “Infrared identification of hard X-ray sources in the Galaxy”, poster by A. Nebot at the X-ray Universe 2014 conference, 16-19th June, Dublin.
- “Infrared identification of hard X-ray sources in the Galaxy”, talk by A. Nebot at the EWASS 2014 : European Week of Astronomy and Space Science to be held on 30 June - 4 July in Geneva, Switzerland.
- “The ARCHES project”, poster by C.Motch at the X-ray Universe 2014 conference, 16-19th June, Dublin.
- “The ARCHES Integrated Cluster Finder”, talk by A. Mints at the X-ray Universe 2014 conference, 16-19th June, Dublin.
- “Testing the unified model of Active Galactic Nuclei in X-ray selected type 1 and type 2 quasars”, talk by S. Mateos Ibanez at the X-ray Universe 2014 conference, 16-19th June, Dublin.
- “Outreach of the European ARCHES project”, talk by L. Michel and A. Schaaf at the IVOA meeting, ESAC, in the education session, Madrid, 19-23 May 2014.

- "MOCs in ARCHES", talk by L. Michel at the VO Interop meeting, held in Hawaii, 27-29 September 2013. It has been focused on the usefulness of VO tools to sort out the catalogue selection.
- "Radio-loud AGN through the eyes of 3XMM, WISE and FIRST/NVS", talk by B. Mingo, 09-2014, Cambridge, UK (New Results in X-ray Astronomy).
- "Breaking AGN sample paradigms with FIRST/NVSS, WISE, and 3XMM", talk by B. Mingo et al. 2015, Llandudno, UK National Astronomy Meeting.
- "The MIXR sample: Radio-loud AGN through the eyes of 3XMM, WISE, and FIRST/NVSS", B. Mingo et al., 2015, "X- ray surveys with advanced multi-wavelength cross-identification methods" (ARCHES scientific conference), Paris, France, 30 Nov.-2 December 2015
- "Discordant optical and X-ray classification of AGN", talk by Ordovás-Pascual et al. 2015, "Highlights of Spanish Astrophysics VIII", Proceedings of the XI Scientific Meeting of the Spanish Astronomical Society held on September 8-12, 2014, in Teruel, Spain, ISBN 978-84-606-8760-3. A. J. Cenarro, F. Figueras, C. Hernández-Monteagudo, J. Trujillo Bueno, and L. Valdivielso (eds.), p. 274-279
- "Exploring XMM-ATLAS with the ARCHES tools", talk by Carrera et al., 2015, "X- ray surveys with advanced multi-wavelength cross-identification methods" (ARCHES scientific conference), Paris, France, 30 Nov.-2 December 2015
- "[OIII] emitting galaxies in the SDSS-dr10/3XMMe/WISE cross-correlation", by Ruiz et al., Spanish X-ray Astronomy 2015, Santander, Spain, 3-5 June 2015
- "[OIII] emitting galaxies in the SDSS/3XMMe/WISE cross-correlation", talk by Ruiz et al., "Demographics and environment of AGN from multi-wavelength surveys", Chania, Greece, 21-24 September 2015
- "[OIII] emitting galaxies in the SDSS/3XMMe/WISE cross-correlation", talk by Ruiz et al., "X- ray surveys with advanced multi-wavelength cross-identification methods" (ARCHES scientific conference), Paris, France, 30 Nov.-2 December 2015
- "The ARCHES Integrated Cluster Finder", talk by A. Schwope & A. Mints, XMM-SSC Consortium meeting #28, MSSL, Jan 2015
- "The integrated Cluster Finder" talk by A. Mints "X- ray surveys with advanced multi-wavelength cross-identification methods" (ARCHES scientific conference), Paris, France, 30 Nov.-2 December 2015

List of ARCHES deliverables

- D2.2 - Report on the kickoff meeting
- D3.1 - Enhanced version of the 3XMM catalogue
- D4.1 - Crosscorrelation tool
- D4.2 - Cross-correlated catalogue
- D5.1 - SED tool
- D5.2 - SED of XMM sources
- D6.1 - Integrated Cluster Finder
- D6.2 - Cluster catalogue
- D7.1 - Archival catalogues: selection criteria and selected list
- D8.1 - Test and validation plan
- D8.2 - Validation reports on ARCHES products
- D9.1 - AGN science case report
- D10.1 - Galactic science case report
- D11.1 - Cluster science case report
- D12.1 - Personnel training report
- D13.1 - Interface to catalogues and SEDs
- D13.2 - Report on dissemination
- D13.3 - Scientific Workshop

1.5 Address of the project public website

www.arches-fp7.eu

Coordinator : Dr. Christian Motch
Observatoire Astronomique
11 rue de l'université
67000 Strasbourg, France
Tel: +33 3 68 85 24 28
Fax: +33 3 68 85 24 32
E-mail: christian.motch@astro.unistra.fr

The ARCHES Consortium (main beneficiaries):

Université de Strasbourg	Observatoire astronomique de Strasbourg
Leibniz-Institut für Astrophysik Potsdam	AIP Leibniz Institute for Astrophysics Potsdam
University of Leicester	University of Leicester
Universidad de Cantabria	IFCA Instituto de Física de Cantabria
Instituto Nacional de Técnica Aeroespacial (INTA)	CENTRO DE ASTROBIOLOGÍA CSIC