DATA ACQUISITION IN THE GENERA CONSORTIUM
Status in 2016

Sveva Avveduto, Daniela Luzi, Ilaria Di Tullio, Lucio Pisacane, Loredana Cerbara, Maria Girolama Caruso, Maria Carolina Brandi

Institute for Research on Population and Social Policies, CNR, Italy

GENERA partners provided information on the availability of institutional and administrative data following the *Statistical data collection template*. The Template provided a specific column on the availability of data, where each organization could insert the label yes/no. This information is relevant to understand the availability of gender related data in the partner organizations with the objective to map the actual gaps and, more importantly, to design a future monitoring system for the implementation of gender policies. Below the list of the different sections of the *Statistical data collection template* are listed and present the picture of the availability of the data in each GENERA partner organization.

Demographics:

Concerning the demographics data, almost all partners have available data on sex, date of birth (or age), citizenship, country of birth of researchers. Concerning the availability of civil status data, only five partners gave a positive answer, while six do not have this information. This means that almost half of the GENERA partner organizations are not able to provide information of wherever the researcher is married, single, cohabitant, widow, widower. etc.
Even more critical is the availability of information on the family work load of the researchers. In particular, only two partners have the number of children and only two partners have data on other familiar components. This information could be relevant to understand the family load and responsibility as well as to design support policy accordingly. Questions of privacy and, in some Countries, even of legal difficulties in asking for, gathering and distribute these kind on information have been claimed by some partners.

Fig. 3.1 Availability of demographics data – Partners distribution by variables
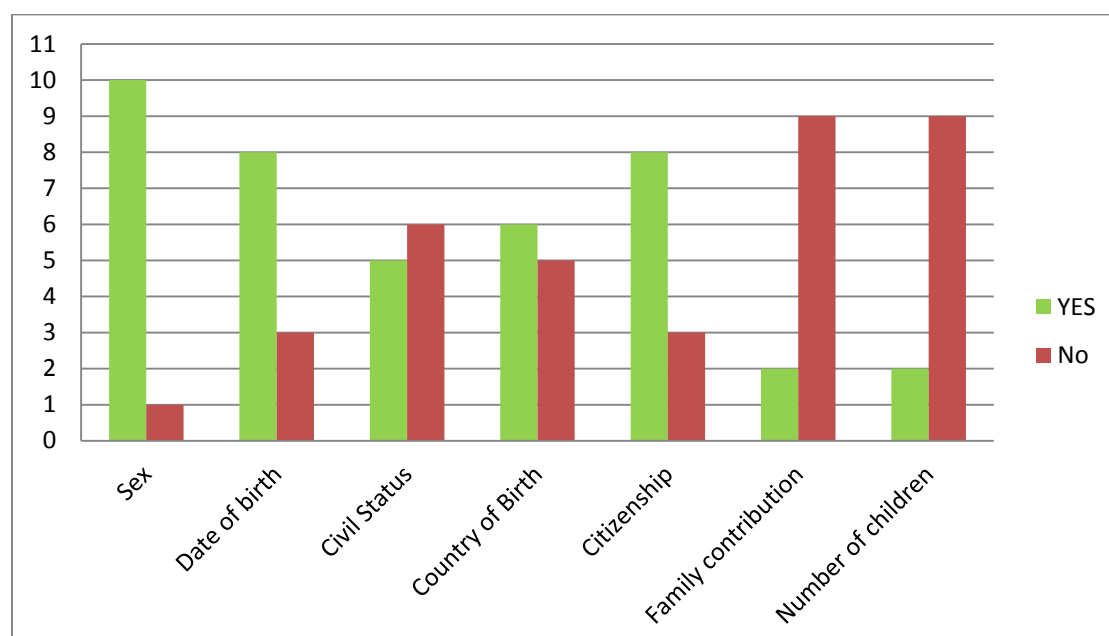
*Tab. 3.1 Availability of demographics data by partner organization*

Demographics data

	Sex	Date of birth	Civil Status	Country of Birth	Citizenship	Family contribution	Number of children
DESY	Y	Y	Y	N	Y	N	N
CNR	Y	Y	Y	Y	Y	Y	N
CNRS	Y	Y	N	N	Y	N	N
UJ	Y	Y	N	Y	Y	N	N
IAC	Y	Y	Y	Y	Y	Y	Y
UNIGE	Y	Y	Y	Y	Y	N	N
KIT	N	N	N	N	N	N	N
IFIN-HH	Y	Y	N	Y	Y	N	N
INFN	Y	N	N	N	N	N	N
FOM	Y	Y	Y	Y	Y	N	Y
MPG	Y	N	N	N	N	N	N

Education and Qualification:

Almost all partners have available data on Master degree and PhD of their researchers . Although the information on PhDs is available, only four partners can collect data on the specific field of degree. Just one partner has available information on the subfield of Master degree. Only three partners have details on the year of the accomplished Master degree and only two on the country where the Master degree was obtained. The information in the subfield of Master degree is therefore missing in almost all institutions, while it is relevant to get a clear understanding of differences among Physics sub disciplines and the career related questions.

Concerning PhD, nine partners can provide this information and of those only three can complement it with the year of degree. Finally only in two cases the country of PhD degree is available. This information is relevant to map the mobility of researchers and can be used to plan relevant policies to support foreign scientists or to attract national researchers trained abroad.

Fig. 3.2 Availability of Education Qualification – Partners distribution by variables
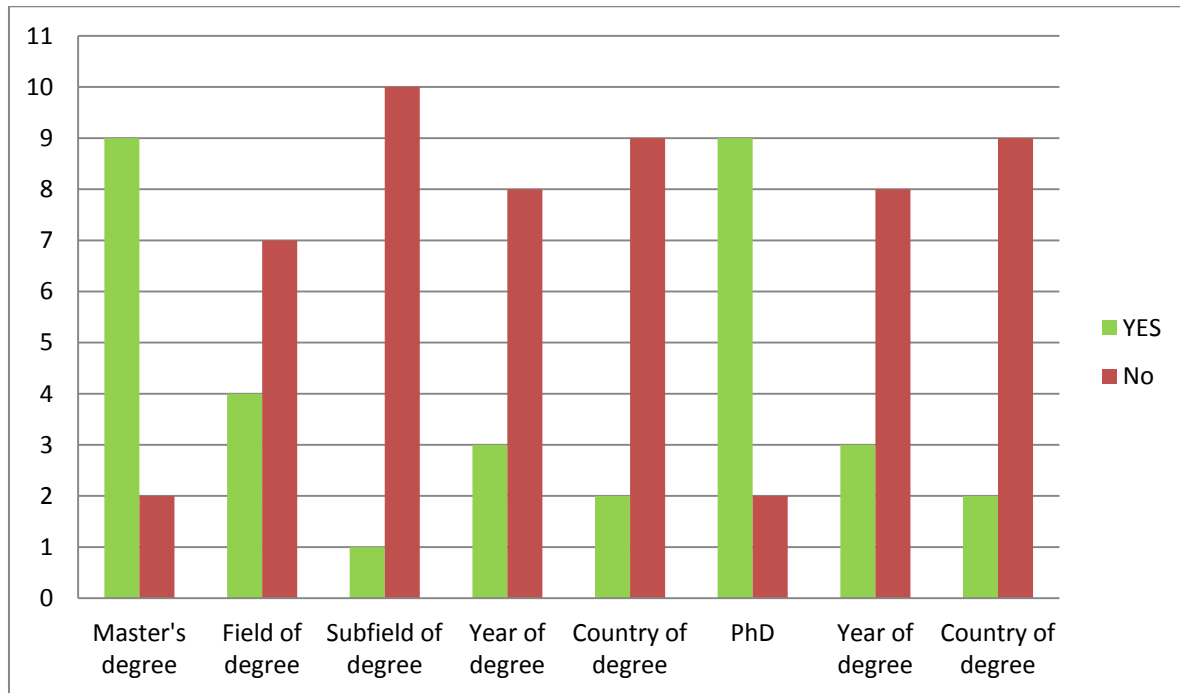
Tab. 3.2 Availability of Education Qualification by partner organization

	Education qualification							
	Master's degree	Field of degree	Subfield of degree	Year of degree	Country of degree	**PhD**	Year of degree	Country of degree
DESY	Y	N	N	N	N	Y	N	N
CNR	Y	Y	N	Y	N	Y	Y	N
CNRS	Y	N	N	N	N	Y	N	N
UJ	Y	Y	N	Y	Y	Y	Y	Y
IAC	Y	Y	Y	Y	Y	Y	Y	Y
UNIGE	N	N	N	N	N	N	N	N
KIT	Y	Y	N	N	N	Y	N	N
IFIN-HH	Y	N	N	N	N	Y	N	N
INFN	Y	N	N	N	N	N	N	N
FOM	Y	N	N	N	N	Y	N	N
MPG	N	N	N	N	N	N	N	N

Career path:

Almost all partners can collect data on the type of the current work contract. Nine partners can provide data on the level of the current contract and just five partners have data on the field of science of the current contract. Eight partners have available information on the starting date of the current contract and seven on the end date. Five institutions can provide data related to the first permanent contract of the researcher and six on other contracts. A lack of information is registered on the data related to the field of Science, to the level of the contract and on the starting and ending Date for all typologies of contracts. Generally the information on the current contract is more precise than data on previous contracts, especially if related to levels, field of science and starting/ending dates. This results in a relevant lack of information to reconstruct the career path of researchers. The interviews to be carried out can partly complement this information and fill some gaps.

Fig. 3.3a Career Path: Availability of Current Contract – Partners distribution by variables
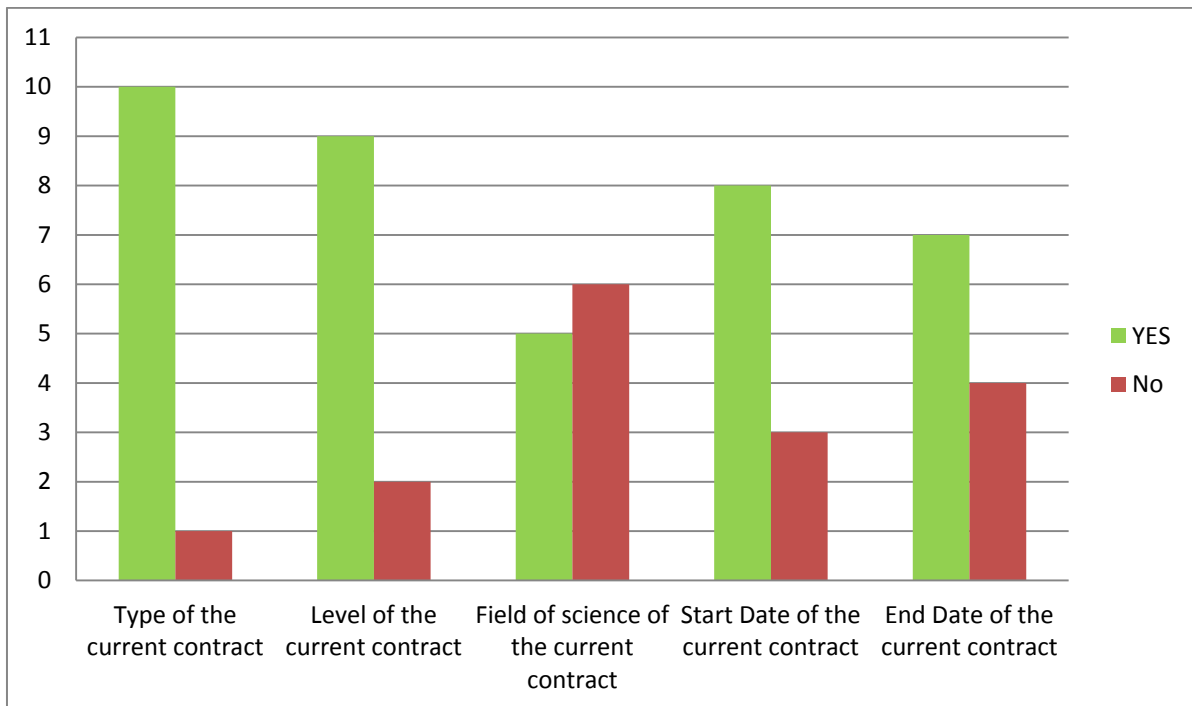
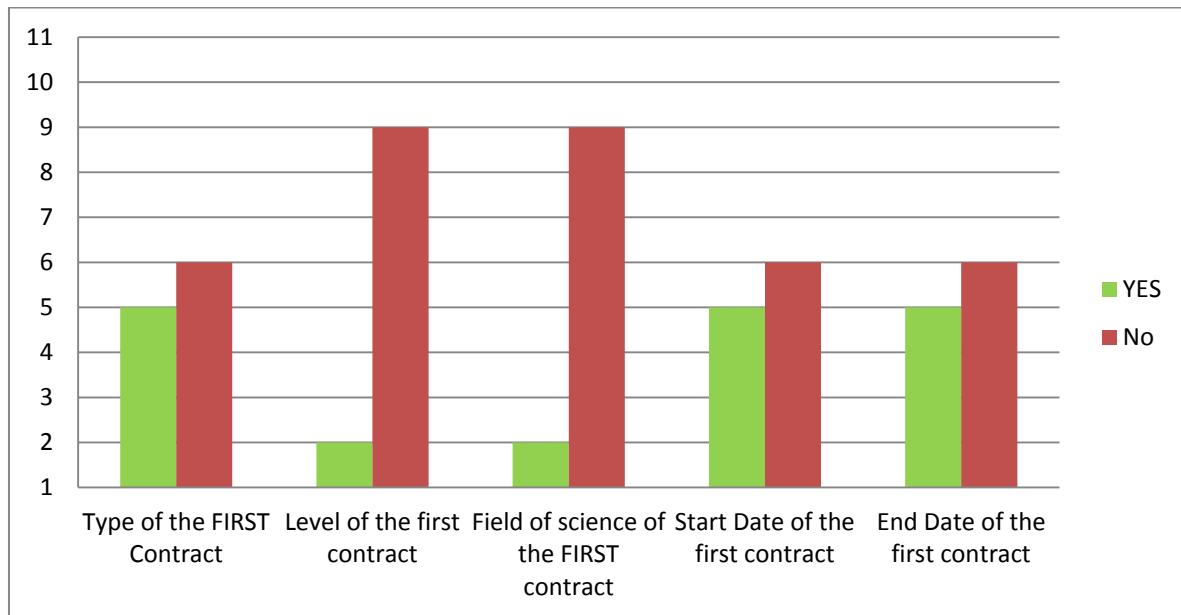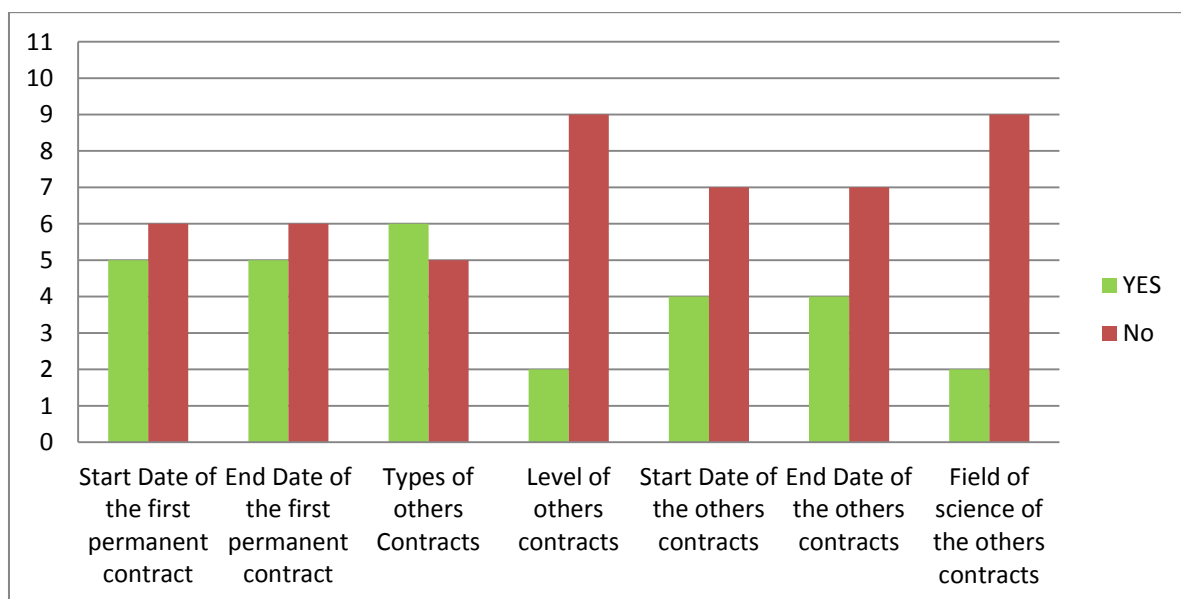*Fig. 3.3b Career Path: First Contract – Partners distribution by variables*

Fig. 3.3c Career Path: Availability of Other Contracts – Partners distribution by variables

Tab. 3a Career Path: Availability of data on current contract by partner organization

	Career Path											
	Type of the current contract	Level of the current contract	Field of science of the current contract	Date of the current contract	start date	end date	Type of the FIRST Contract	Level of the first contract	Field of science of the FIRST contract	Date of the first contract	start date	end date
DESY	Y	Y	N	Y	Y	Y	N	N	N	Y	Y	Y
CNR	Y	Y	Y	Y	Y	N	Y	Y	N	Y	Y	CY
CNRS	Y	Y	Y	N	N	N	N	N	N	N	N	N
UJ	Y	N	N	Y	Y	Y	Y	N	N	Y	Y	UJ Y
IAC	Y	Y	Y	Y	Y	Y	N	N	N	N	N	IA N
UNIGE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
KIT	Y	N	Y	N	N	N	Y	N	Y	N	N	KI N
IFIN-HH	Y	Y	N	Y	Y	Y	N	N	N	N	N	IF N
INFN	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	IN Y
FOM	Y	Y	N	Y	Y	Y	Y	N	N	Y	Y	FO Y
MPG	N	Y	N	N	N	N	N	N	N	N	N	M N

Tab. 3b Career Path: Availability of data by partner organization

	Career Path								
	Date of the first permanent contract	*start date*	*end date*	*Types of others Contracts*	*Level of others contracts*	*Date of the others contracts*	*start date*	*end date*	*Field of science of the others contracts*
DESY	Y	Y	Y	Y	N	N	N	N	N
CNR	Y	Y	Y	Y	Y	Y	Y	Y	N
CNRS	N	N	N	N	N	N	N	N	N
UJ	Y	Y	Y	Y	N	Y	Y	Y	N
IAC	N	N	N	N	N	N	N	N	N
UNIGE	Y	Y	Y	Y	Y	Y	Y	Y	Y
KIT	N	N	N	Y	N	N	N	N	Y
IFIN-HH	N	N	N	N	N	N	N	N	N
INFN	Y	Y	Y	N	N	N	N	N	Y
FOM	Y	Y	Y	Y	N	Y	Y	Y	N
MPG	N	N	N	N	N	N	N	N	N

Work organization

Six partners have data on researchers' responsibility attaining management of institutes and/or research structures. Only three organizations can provide information on research project responsibility, four on laboratories management and four on teaching activities. A general lack of information is registered on the duration, starting date and ending date of those responsibilities. Five partners have available data on part-time work and parental leave. Only four institutions have data attaining careers breaks. The information on the duration of part time, parental leave and careers break is partially present, even where the activity is registered. This results in a partial availability of data on work organization, where activities such as parental leave, teaching, careers break are monitored.

Fig. 3.4a Work Organizations: Availability of data on Responsibility – Partners distribution by variables
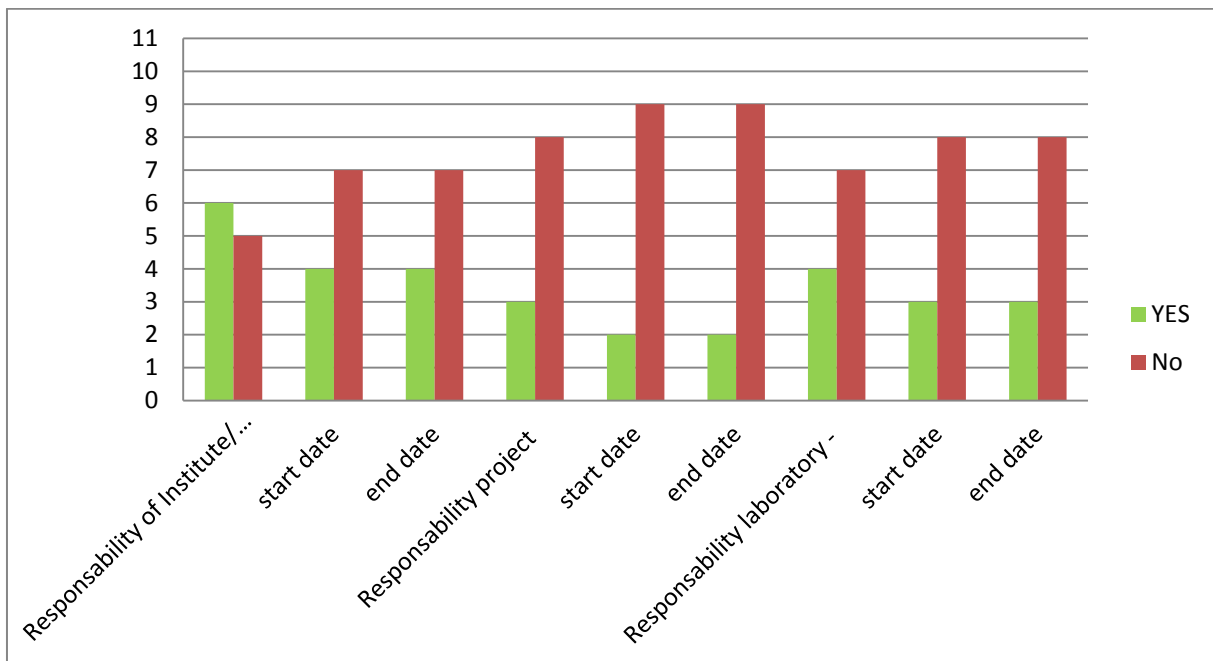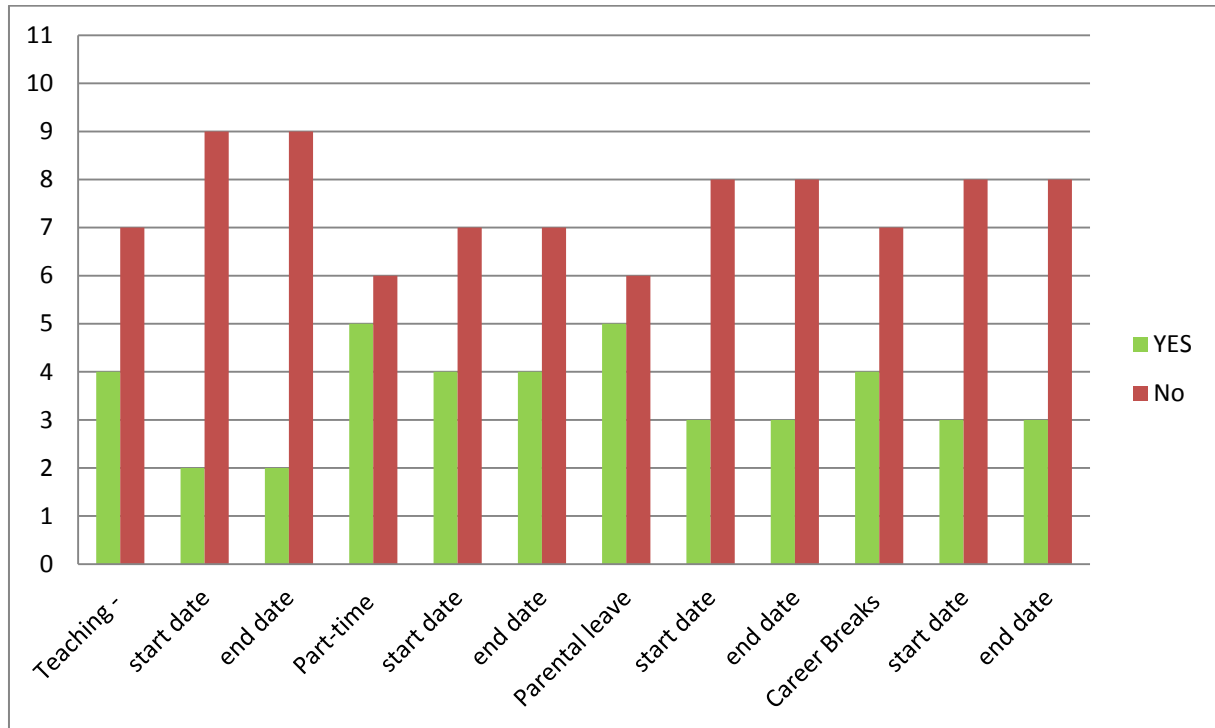
Fig. 3.4b Work Organizations: Availability of data on Leaving Periods – Partners distribution by variables

Tab. 3.4a Work Organization: Availability of data by partner organization

Work organization									
	Responsability of Institute/ Structure etc.	*start date*	*end date*	*Responsability project -*	*start date*	*end date*	*Responsability laboratory -*	*start date*	*end date*
DESY	N	N	N	N	N	N	N	N	N
CNR	Y	Y	Y	Y	N	N	N	N	N
CNRS	Y	N	N	N	N	N	Y	N	N
UJ	Y	Y	Y	Y	Y	Y	Y	Y	Y
IAC	Y	Y	Y	Y	Y	Y	Y	Y	Y
UNIGE	N	N	N	N	N	N	N	N	N
KIT	N	N	N	N	N	N	N	N	N
IFIN-HH	Y	N	N	N	N	N	N	N	N
INFN	Y	Y	Y	N	N	N	Y	Y	Y
FOM	N	N	N	N	N	N	N	N	N
MPG	N	N	N	N	N	N	N	N	N

Tab. 3.4b Work Organization: Availability of data by partner organization

	Work organization											
	Teaching -	*start date*	*end date*	**Part-time**	*start date*	*end date*	**Parental leave**	*start date*	*end date*	**Career Breaks**	*start date*	*end date*
DESY	N	N	N	N	N	N	N	N	N	N	N	N
CNR	N	N	N	N	N	N	Y	N	N	Y	N	CN
CNRS	Y	N	N	Y	N	N	Y	N	N	N	N	CN
UJ	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	UJ
IAC	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	IY
UNIGE	N	N	N	N	N	N	N	N	N	N	N	U
KIT	N	N	N	N	N	N	N	N	N	N	N	KIT
IFIN-HH	Y	N	N	N	N	N	N	N	N	N	N	N
INFN	N	N	N	Y	Y	Y	N	N	N	N	N	INFN
FOM	N	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y
MPG	N	N	N	N	N	N	N	N	N	N	N	MN

Research output:

The situation about the research output reveals a lack of data in almost all GENERA partners. Only three partners have data on research outputs related to journal article, conference contribution, chapter in edited books, report, book, patent/trademark.

This information may be available in other archives such as institutional repositories as some partners pointed out in the questionnaire. It could however be difficult to link the data with the administrative ones.

Fig. 3. Research Output: Availability of data Partners distribution by variables
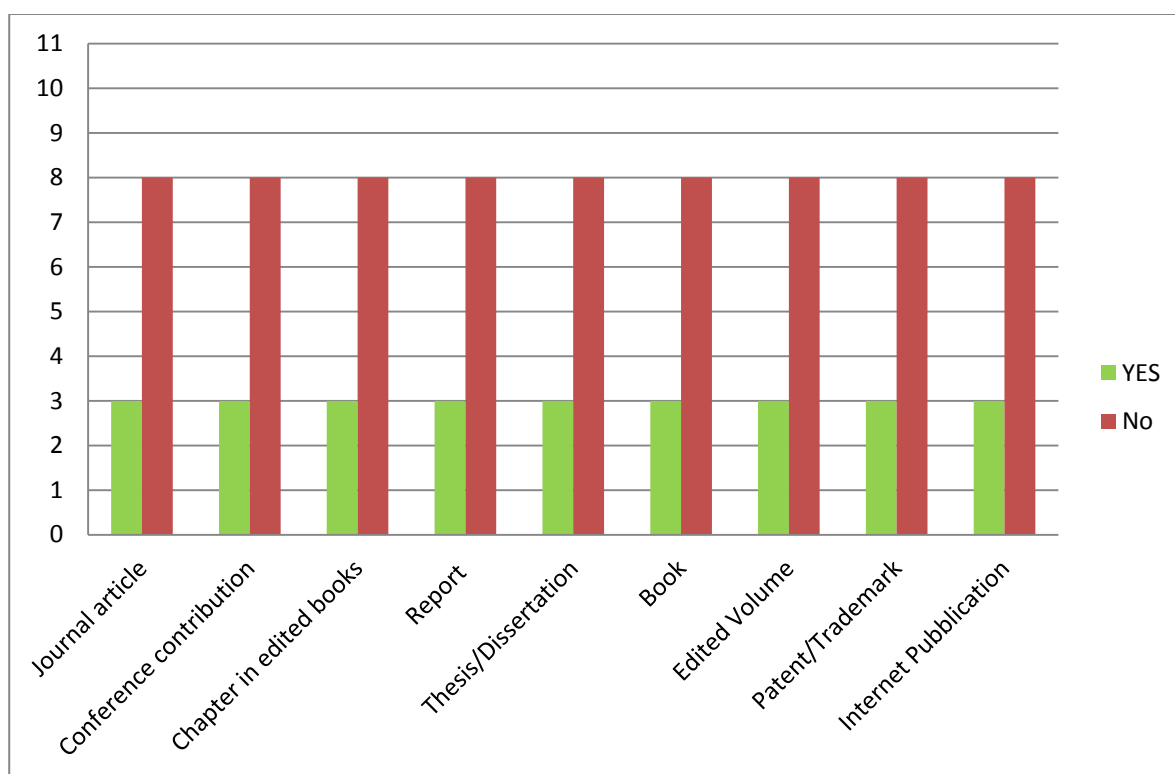
Tab. 3.5 Research Output: Availability of data by partner organization

	Research Output								
	Journal article	Conference contribution	Chapter in edited books	Report	Thesis/ Dissertation	Book	Edited Volume	Patent/ Trademark	Internet Publication
DESY	N	N	N	N	N	N	N	N	N
CNR	Y	Y	Y	Y	Y	Y	Y	Y	Y
CNRS	N	N	N	N	N	N	N	N	N
UJ	N	N	N	N	N	N	N	N	N
IAC	Y	Y	Y	Y	Y	Y	Y	Y	Y

UNIGE	N	N	N	N	N	N	N	N	N
KIT	Y	Y	Y	Y	Y	Y	Y	Y	Y
IFIN-HH	N	N	N	N	N	N	N	N	N
INFN	N	N	N	N	N	N	N	N	N
FOM	N	N	N	N	N	N	N	N	N
MPG	N	N	N	N	N	N	N	N	N

The availability of institutional and administrative data within partners' organizations present an overall lack of gender relevant data in almost all sections of the *Statistical data collection template,* excluding very few such as, for instance, demographics and current work contract*.*

The development of *ad hoc* surveys to enrich the data collection can be an important starting point to improve the monitoring activities. As administrative offices can gather not all the data, as it may be sensitive or legally impossible, there is a need also to conduct studies within institutions that will support programmes and policies e.g. study on career paths. So some data could be gathered in a way that it is not available to institutions as far as data on individuals is concerned, but it provides elaborated aggregated data results on researchers' situation and career. Researchers might not be willing also to share some specific data with an institution administration e.g. family situation.

This result, even if negative, is to be considered important for the GENERA project as it helps to understand and to foresee the mapping of data and information in research organizations, starting from the analysis of the current status of data availability.

This gives rise to a call for action for HR managers, central administration offices and, in general, for the management of the organizations, to establish a proper data collection that includes gender relevant issues. A rich and detailed database can support a gender plan and an equality monitoring system for gender policy that can prepare measures that can have a real impact on researcher's lives and work.

This suggests the need of strengthening the process of gathering and measuring information on targeted variables in an established system. This might enable us to answer relevant questions and evaluate outcomes. For these reasons it is important to enhance Gender equality offices role and structures with the objective of coordinating and facilitating the collection of the data in order to establish a long term monitoring system for research organizations.

SECTION 4: OUTLINE OF DATA ACCESS IN VIEW OF FUTURE MONITORING ACTIVITIES

Following the setting up of the Data collection Framework and the consequent availability of gender relevant data among partners organization, the WP2 team decided to focus on the analysis of hindering and facilitating factors in the data collection process. With these objectives a questionnaire "*Outline of data access in view of future monitoring activities*" was sent out to the GENERA partners to collect opinions on the different issues faced during the internal process to access secondary administrative data in each organization. The questionnaire had both qualitative list of items (with a scale of value) and open questions where partner were free to insert their comments in narrative form.

The questionnaire was presented during the project Joint Sectariat (July 2016, DESY, Hamburg) where it was discussed in a plenary session. At end of September all partners (10), with the exception of Horia Hulubei National Institute of Physics and Nuclear Engineering (IFIN-HH), answered the questionnaire.

The data collection process, as discussed in the previous section of this report, has highlighted consistent gaps in data availability and/or incompleteness of relevant information in many partner organizations. The questionnaire gave partners the possibility to explain the reasons of those gaps, as well as to specify that data may be registered but not available or not accessible for monitoring activities. In this sense many comments of the questionnaire underlined that data availability in research organizations is also a policy and management issues. Without a clear policy aiming to set up a monitoring system on gender policy and strategy the data will be always difficult to access due to lengthy bureaucratic procedures, fragmentation or incompleteness of institutional databases.

The questionnaire was composed of four parts:

- Factors that facilitated the collection of gender relevant data;
- Aspects that hindered the collection of gender relevant data;
- Positive actions to improve gender relevant data monitoring activities;
- Further general comments.

The first three parts presented a number of options to be selected by respondents with additional open questions for comments. The fourth section was an open question with narrative answers. The answers and the results of the questionnaire are presented below. A full version of the submitted questionnaire is reported in the Annex 4.

Section 1: Factors that facilitated the collection of gender relevant data

Options of the questionnaire :

- Availability of central management databases

- Single contact point to obtain data from administrative source

- Presence of Gender equality office or similar dedicated structures

- Already developed GEPs

- Codified procedure to obtain relevant statistical data on personnel and scientific production

The answers provided by partners converge on three closely related facilitation factors the "*Availability of central management databases*" and the '*Presence of Gender equality office or similar dedicated structures*' and a '*Single contact point to obtain data from administrative source*' pointing out that a dedicated structure and the development of an infrastructure are necessary elements to facilitate data collection of the gender relevant data.

Worth noticing is that the availability of GEP is considered a facilitating factor by a consistent number of partners even if it is not considered among the most important ones. This can be explained by some comments that consider the implementation of GEPs as a starting point for setting up procedures and policies for the collection of data.

As a partner commented summing up facilitating factors: '*Central management databases are very important but it's also very important that these data can be available for staff involved in Gender Equality for diagnosis and monitoring of Gender Equality policies in the institution. A codified procedure to obtain data is certainly useful especially in big organizations.*'

Section 2: factors that hindered the collection of gender relevant data

Options of the questionnaire:

- Institutional databases fragmented over different offices and administrative branches

- Databases managed by different offices requiring more than one procedure to obtain access

- Data analysis requiring heavy data recodification and cleaning

- Presence of duplicated record-data

- Privacy issues

- Data non covering long period that enable the reconstruction of careers paths

- Administrative database available on proprietary software, often not easily compatible with statistical tools

- No codified procedure to obtain relevant statistical data on personnel and scientific production

The most perceived hindering factor in data collection process has been privacy issues. Privacy has been raised several times by different administrative offices as a obstacle to cooperate in giving out data on

personnel. Some organization, even if data are anonymized, fear that it might be possible to identify the individual data, thus not respecting privacy.

As stated by one partner reflecting the experience of many other ones:

'The privacy issue has been the most relevant one for us: the offices refused to provide data on individual basis, arguing that, even if anonymised, it would not respect privacy'.

Although the answers present an high variability, some missing structural factors are mentioned recurrently especially in the comments:

'There is not a central institutional database; data are not collected and stored centrally. They are fragmented overall several offices and thus, they are managed by more than an office, so the information given by a certain data can be incomplete. It is also difficult to obtain the data, because of many procedures to gain access to them'.

'Fragmentation of databases is not necessarily bad if a coordination between departments and Gender Equality Commission or Office exists. Cleaning of data is always complicated: it's better to have some previous standard or procedure for the maintenance of databases from which data are obtained'.

"The data available at (name of the organization) for employees and non-permanent staff are incomplete since the career path is not followed. Some data, for instance on publications, would be recoverable from other databases. Also Privacy is an issue for information for instance on benefits concerning children".

Section 3: positive action to improve gender relevant data monitoring activities

Options of the questionnaire:

- Collection and updating of a minimum dataset to monitor gender relevant activities

- Set up a codified procedure to request relevant data

- Relevant anonymized administrative data available in open access

- Include in the collection of data form non-permanent staff (research, grants, fellows, internship) with the aim of monitoring early career

The actions that could positively improve data monitoring activities in research organizations have been identified mainly in setting up a specific collection, updating and monitoring procedures of gender related data, closely linked to the relevance of including to database also all data on temporary personnel, as these researchers are usually most affected by gender bias. Moreover, this can allow to analyze the trajectories of early stage researchers' career.

Section 4: Further general comments

This section proposes an open field where partners could provide narrative comments and inputs on the main issues faced during the data collection process in their organizations. The section gathered many diversified comments related to different aspects. These comments, together with those expressed in the questionnaire *"Outline of data access in view of future monitoring activities"* have been gathered and represented below by sub topic clusters which mostly emerged from the GENERA partners' comments.

The Organizational Level

At the organizational level the major problems emerging from the questionnaire refer to the complex structure of the organizations. For instance the presence of different institutes and departments may determine separate data collection that are consequently more difficult to gather and harmonize at central level.

A contact point/office or system that could collect and provide statistical data at central level is generally lacking in many organizations, due also to their intrinsic structure. As shown by a partner's comment "*it is important to have a gender equality office/personnel department that collect them and that is responsible for them*". Another relevant problem is related to the incompleteness of the data, in fact they are mostly demographics data and it is difficult to reconstruct the career path as reported by a partner "*the absence of a central institutional database constituted the first real problem that hindered the collection of gender relevant data. In fact data are not collected at a central level; only some kind of data are collected by several and different offices, so they are very incomplete and fragmented*".

The main suggestions emerged by partners refer to the possibility to have a:

- Coordination between departments and statistical management office

- Early involvement of management

- Already developed GEP

- A central gender equality office

a. The Statistical Level

Strictly related to the Organizational aspects, the main issues at statistical level concern the methodology of data collection.
They are related especially to the fragmented and spread databases due to the different way to collect data. In fact partners referred that even if in the majority of the organizations the statistical data are not available, they may not be collected at central level but may be available at local decentralized level. As reported by a partner: "*the data available for employees and non-permanent staff are incomplete since the career path is not followed. Some data, for instance on publications, would be recoverable from other databases*". Another partner added: "*the data currently available are mostly demographic data and do not allow to study e.g. career paths and career breaks*".

The main suggestions emerged by partners refer to the possibility to have a:

- Unique central statistical management office

- Unique central institutional database on personnel

- Codified procedure to obtain data

- Completeness of the data

b. The Privacy issue

Privacy represents an important issue. In fact many partners suggested that some data are not gathered because of privacy rules, as stated before. As shown by a partner "*even if some data were available, they could not be easily used in the analysis because of privacy laws, they could only be used in an anonymous way*". An even more difficult situation is present in those organizations which consist of different small institutes because, as stated by a partner: "*seeing that the number of persons from one sex/occupation/national background can be rather small in the institutes, administrative data can never be fully anonymized and thus cannot be available in open access, due to the rigid and thorough data protection and ethics standards*", another partner adds: "*Privacy was a highly debated issue (..). To gain access to the rest of them (including sex, date of birth, year of degree, information about leaving periods), we need the consent of each individual employee. Unavailability of these data – especially of the sex variable – hinders the analysis*"

The main suggestions emerged by partners refer to the possibility to have a:

- Clear procedures to obtain data in organizations

- Codified procedures to anonymize data in organizations

Improvement suggested

Summing up the general comments the following suggestions emerged:

- Set up a minimum set of data

- Fix a gender monitoring system

- Fix a recurring refreshing of the data

- Changing attitude towards gender equality

- Raising awareness of policy makers

- Promote the implementation of gender equality plan