CRD2201/02: Data Skills and Data Assets

October 2022

SAPIO RESEARCH

Disclaimer

This report (the “Project”) has been prepared by Sapio Research Limited (“Sapio Research”) for the WDNA (the “User”). The defined term “Sapio Research” shall mean Sapio and its affiliates, and each of their former, current or future owners, partners, members, directors, managers, officers, directors, employees, attorneys and agents and the successors and assigns of the foregoing persons. Sapio Research reserves the right to amend, supplement or replace this report at any time. User shall not rely on any oral communications by Sapio Research employees or representatives with respect to the Project, and the opinions, projections, estimates and conclusions of Sapio Research are solely those set forth in and qualified by this report.

The sole purpose of this report is to assist the User in evaluating the Project and this report shall not be used for any other purpose. User acknowledges that it has, either alone or in conjunction with its advisors, made an independent investigation into the advisability of the Project and that this report is not the sole basis for User’s ultimate course of action with respect to the Project. Sapio Research is not and shall not be responsible for decisions made by User.

User may not distribute, reproduce, or disclose this report, in whole or in part, to any third party unless and until User receives the prior written consent of Sapio Research.

This report is based on information available at the time this report was prepared and on certain assumptions, including, but not limited to, assumptions regarding future events, developments and uncertainties and contains “forward-looking statements” (statements that may include, without limitation, statements about projected revenues, earnings, market opportunities, strategies, competition, expected activities and expenditures, and at times may be identified by the use of words such as “may”, “could”, “should”, “would”, “project”, “believe”, “anticipate”, “expect”, “plan”, “estimate”, “forecast”, “potential”, “intend”, “continue” and variations of these words or comparable words).

Sapio Research is not able to predict future events, developments and uncertainties. Consequently, any of the forward-looking statements contained in this report may prove to be incorrect or incomplete, and actual results could differ materially from those projected or estimated in this report. Sapio Research undertakes no obligation to update any forward-looking statements for revisions or changes after the date of this report and Sapio Research makes no representation or warranty that any of the projections or estimates in this report will be realised. Nothing contained herein is, or should be relied upon as, a promise or representation as to the future.

Sapio Research has conducted the Project with due care and in accordance with generally accepted standards in the research industry, however it does not guarantee that the data collected, processed, analysed and provided is complete and accurate in all respects.

If you are an authorized recipient of this report (other than User), then you have already signed an agreement whereby you irrevocably agreed to fully release and hold harmless Sapio Research, its affiliates and their representatives. Such agreement remains in full force and effect.

Background to project

Project background

Project background

- The WDNA (Welsh Data Nation Accelerator), came to us with the task of conducting scoping reports on Data Assets and Data Skills.
- The WDNA is a pan-Wales initiative that is being developed to accelerate new insight, foresight and intelligence from diverse data assets for societal, health and economic impact.
- The goal of the WDNA is to bring together data assets and skills to facilitate innovation.
- The research focused on uncovering what **data assets** were available along with how **data skills** are currently being developed.
- All information would, ideally, be specific to Wales and come under the sectors of Health and Wellbeing, Public Services and Innovation, Net Zero and the Environment, Creative and Professional Services, and Future Manufacturing and Systems

Desired deliverables

- Locate the **data assets** currently available in Wales and list the key details about each asset (e.g., open-source or paid, structured or unstructured etc.)
- Collate a list of **data skills** by examining what training is being offered in Wales currently
- Provide a breakdown of each course to illuminate the skills it develops
- Offer a final analysis on the gaps in knowledge for both data assets and data skills
- It is important to note that we were not attempting to learn how Wales compared to other places in the UK, but rather the provisions it has

Report methodology

- The desk research collected a series of data assets and skills courses (the whole collection can be found on the separate Excel document).
- Using the number of sources and the filters to the columns on the Excel document, we have created quantitative visualisations of the information collected.
- This is designed to provide an overview of the type of information that can be found in the Excel document which can then be used to do a deep dive on the various skills and assets.
- 4 in-depth interviews (IDIs) were conducted, and the information provided by the interviewees has been distilled down into the main takeaways from the interviews.
- These summaries are accompanied by quotations that correlate with the key takeaways from the interviews.
- This report covers both scoping reports, and the information has been harmonized given that these two topics are interrelated

Research Process

Iterative process, with check-ins, not necessarily linear in progression

Immersion Call with Stakeholders

Understand more about the project, areas of interest, challenges, existing insight and intelligence, results from previous research, etc.

Stage 1

Exploratory Desk research

Transpose questions from brief into a research framework.
Identify published information, undertaking desk research to judge quality, *granularity* and *feasibility* of what's achievable

Stage 2

Qualitative research

30-45 minute in-depth interviews with key influencers & industry experts, to test hypothesis, form case studies and add aspects and risks to consider

Stage 3

Validation Desk research

Further desk research to validate new aspects raised during the qualitative work

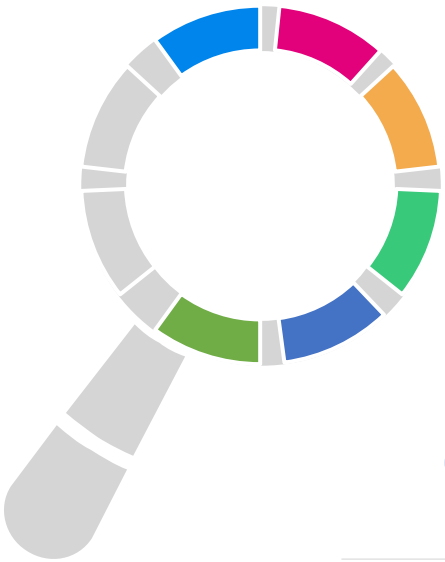
In-depth Interviews methodology

- 4 In-depth interviews were conducted with key industry experts to form case studies, help formulate hypotheses and back up desk research findings
- Contacts were approached with the help of the WDNA
- Interviewees came from the following sectors*: manufacturing, fintech and academia (Creative industries & Health & Wellbeing) and were introduced to Sapio Research by stakeholders of this
- The interviews were conducted via online video calls by Sapio Research in October 2022

*While many contacts were approached to take part, time constraints meant participants were selected on a first come first serve basis

Questions to be answered

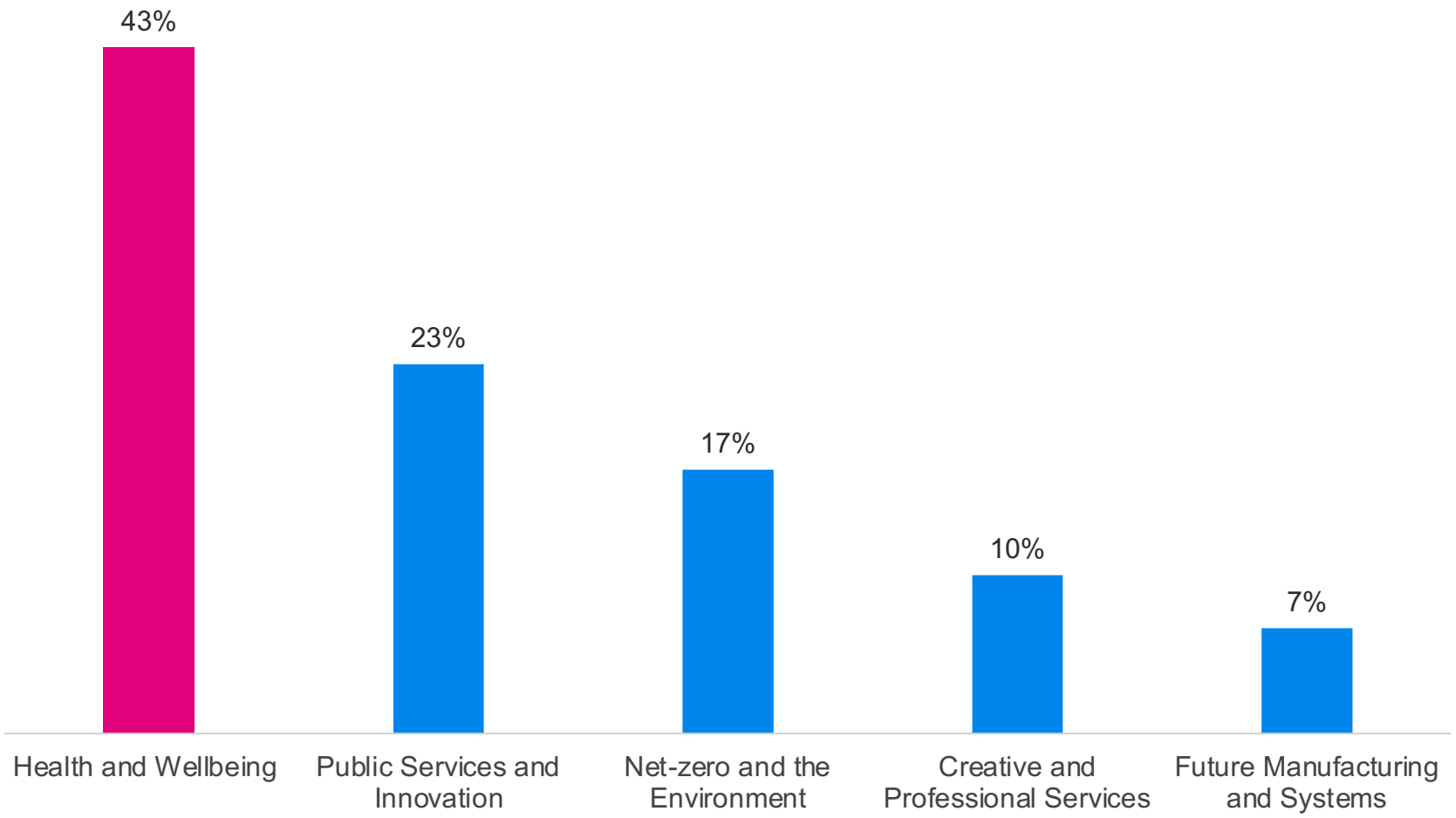
Based on the information supplied by the WDNA we proposed a desk study that addressed the following questions:



- What data assets are currently available, are they specific to Wales and are they free to access?
- Which areas are lacking data assets?
- How do these sources relate to the key themes of the WDNA?
- What course are available in Wales to serve specific data skills?
- What type of person are these courses suitable for?
- Based on the skills identified, what the gaps in data skills in Wales?

Data Assets Findings

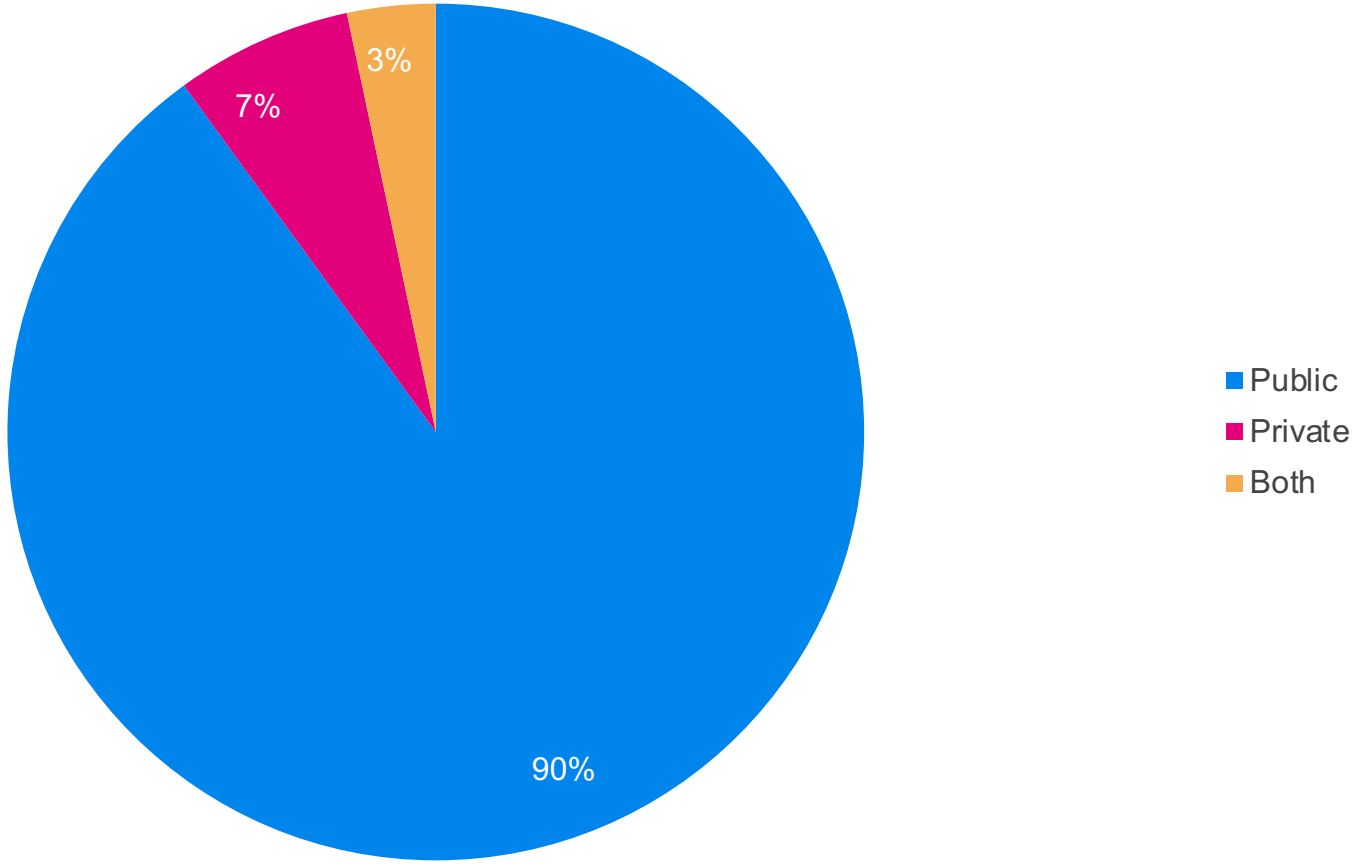
Of the WDNA themes, the data assets came mostly under the Health and Wellbeing umbrella



Base: 30

N.B The numbers provided here have been calculated using the base number of assets (30) and filtering the entries to provide more granular breakdowns.
** See appendix 1 for definition of terms.*

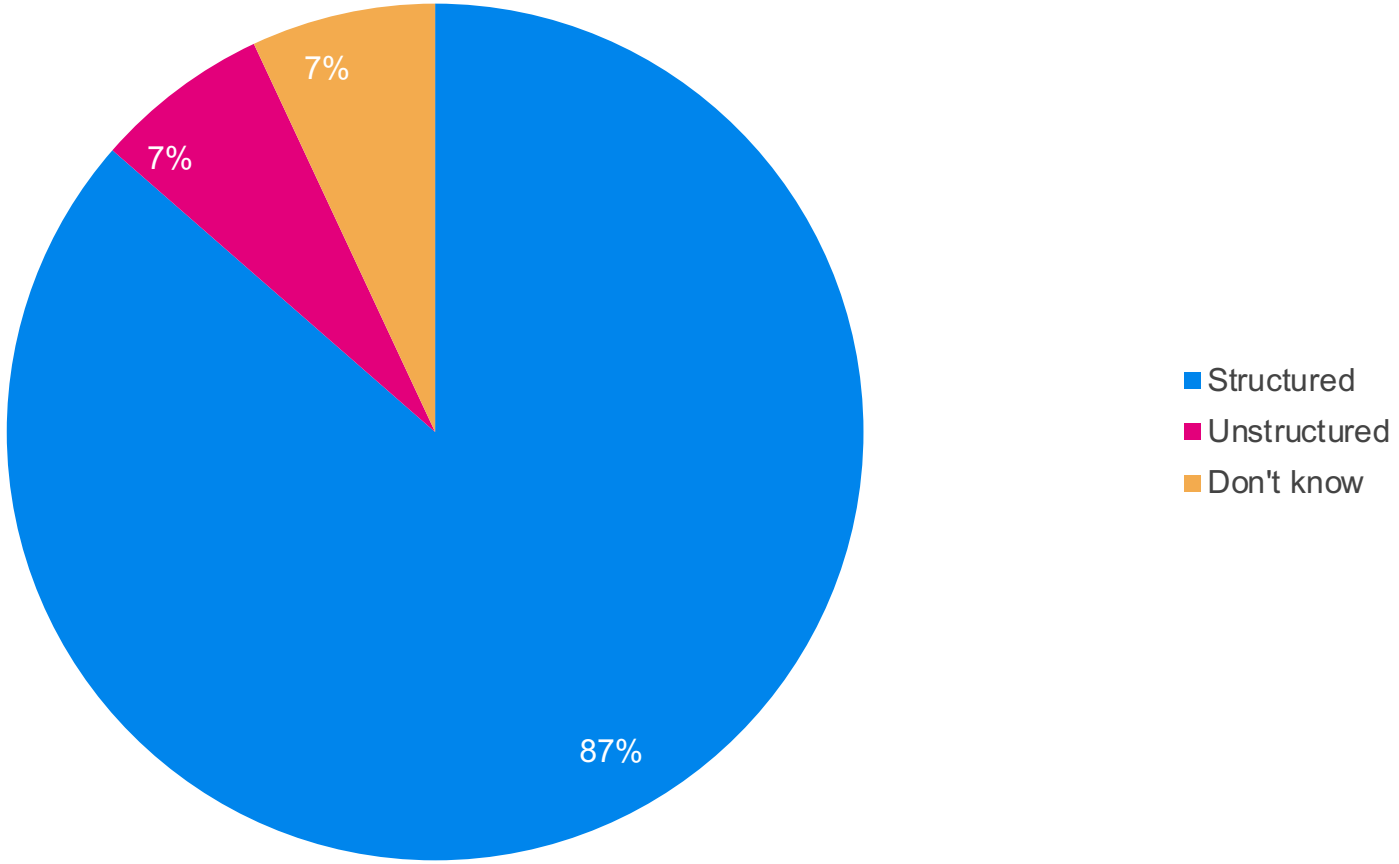
90% of the data assets collected contain data related to the public sector



Base: 30

N.B The numbers provided here have been calculated using the base number of assets (30) and filtering the entries to provide more granular breakdowns.

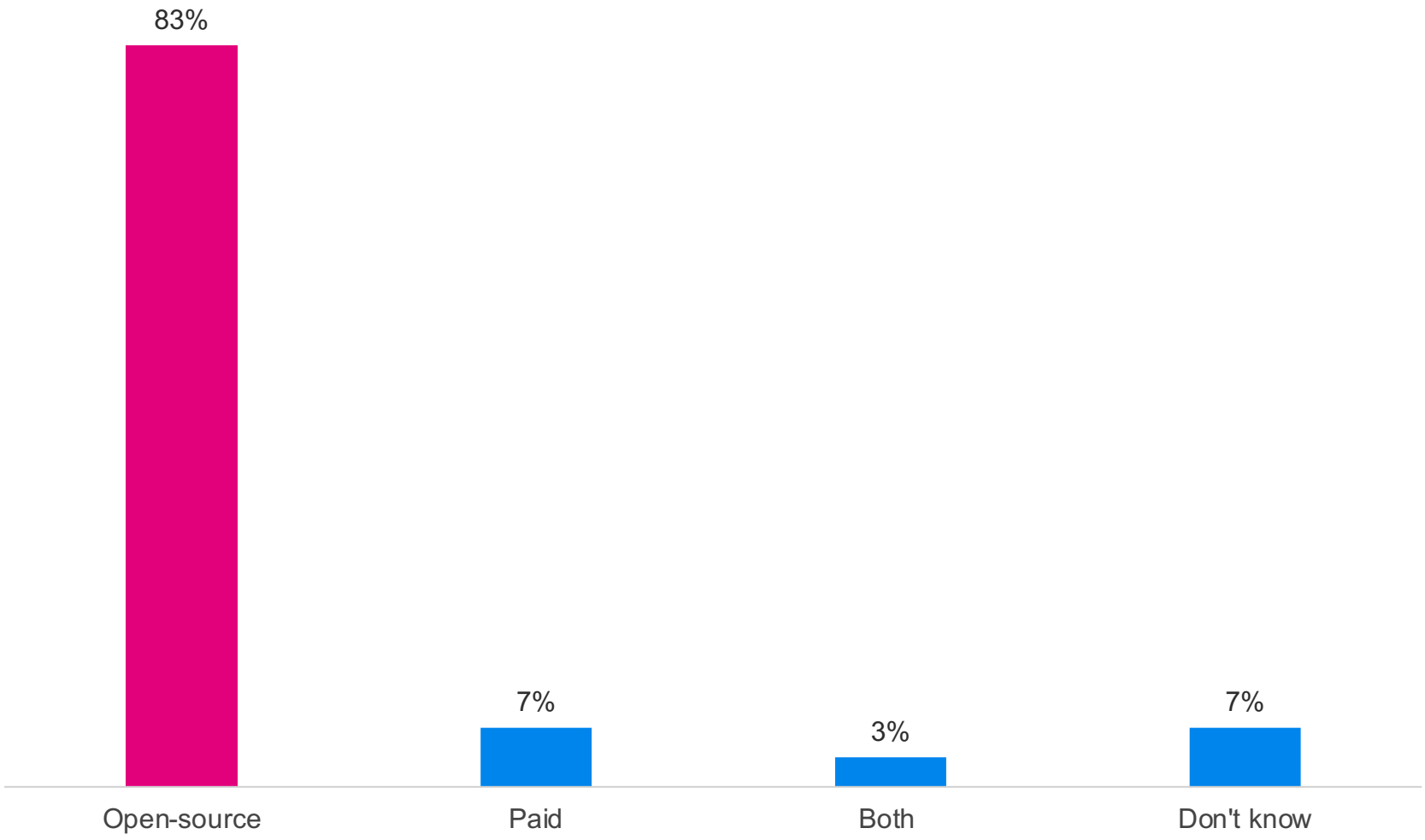
87% of the assets contain structured data, meaning that it does not require cleaning to extract the information



Base: 30

N.B The numbers provided here have been calculated using the base number of assets (30) and filtering the entries to provide more granular breakdowns.

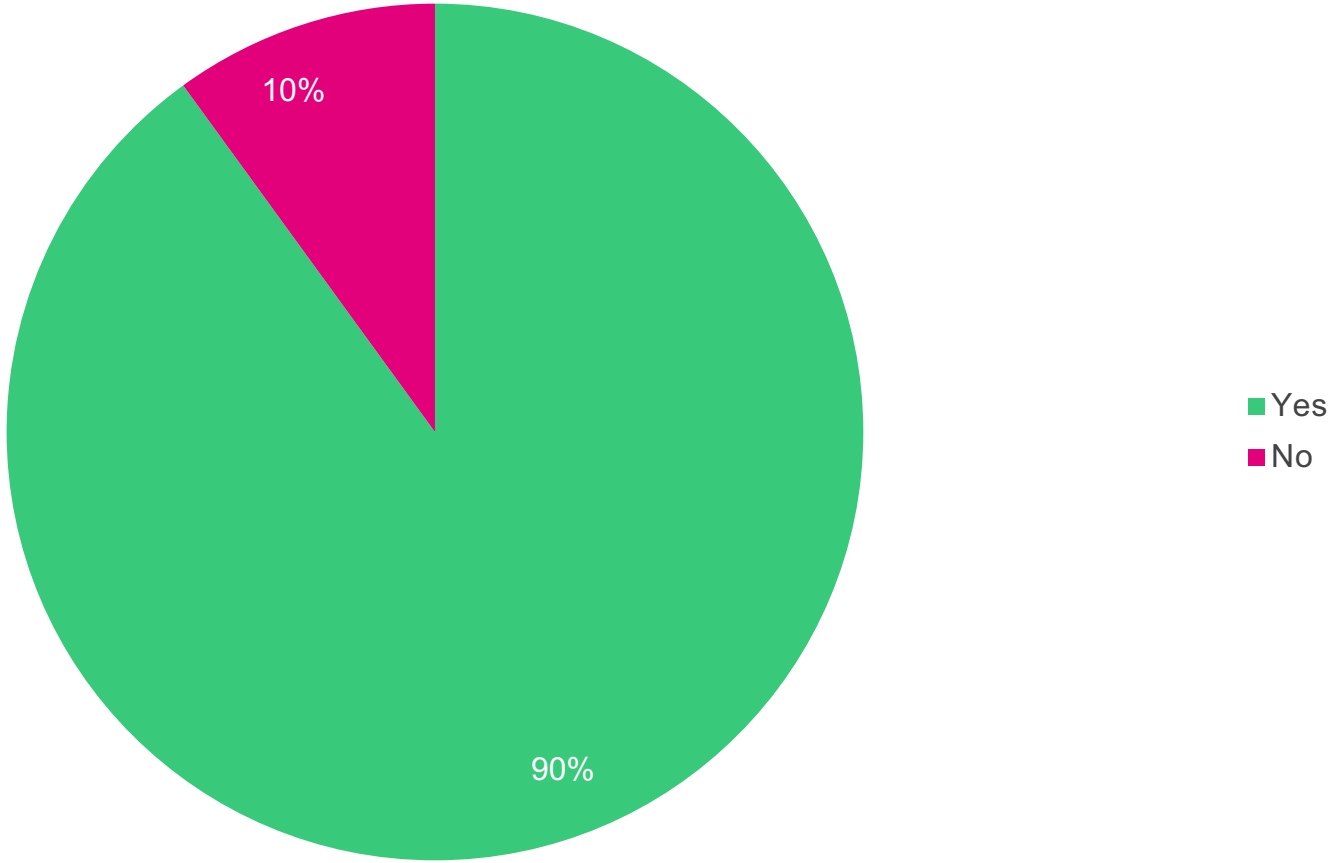
83% of the assets were open-source and can therefore be accessed free of charge



Base: 30

N.B The numbers provided here have been calculated using the base number of assets (30) and filtering the entries to provide more granular breakdowns.

90% of the assets contain data that is specifically about Wales



Base: 30

N.B The numbers provided here have been calculated using the base number of assets (30) and filtering the entries to provide more granular breakdowns.

Summary

- **Assets are not evenly distributed between themes:** In terms of the WDNA themes and the assets we found, the largest collection of assets came under the Health and Wellbeing banner, whilst there was a scarcity of assets for Future Manufacturing and Systems. To that end, efforts should be made to locate further assets that can provide data on Future Manufacturing and Systems.

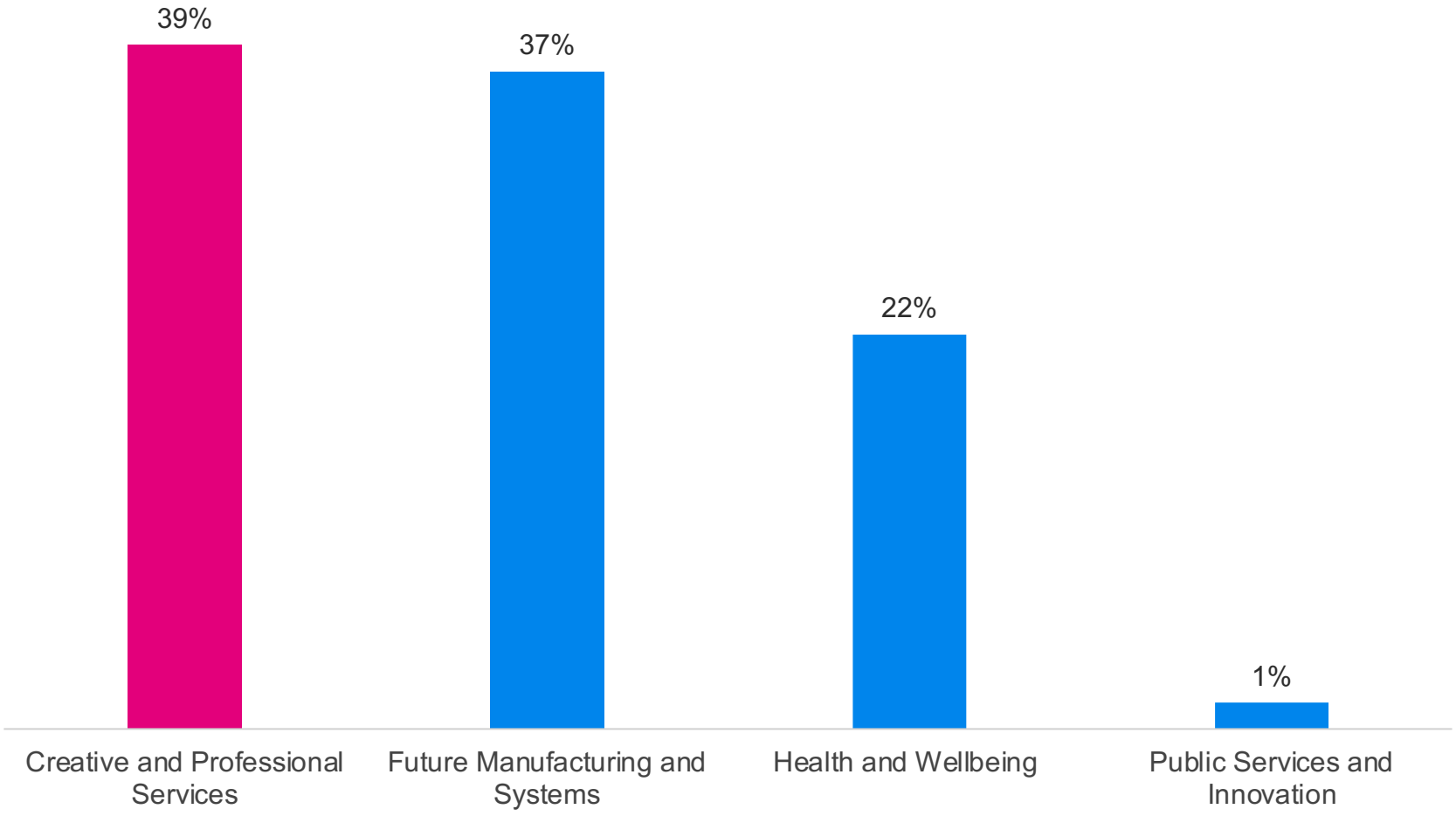
- **Public assets more prevalent than private:** We were mainly able to locate public rather than private sources as private companies do not advertise the availability of their data. That said, there are a significant number of public assets (27) and 25 of the assets collected are free to access. This should not, however, be taken as a statement of private assets not having data. Rather, we were unable to confirm that private companies had data that was publicly available.

- **There is a great deal of data on Wales:** In addition to there being a large number public assets, 27 of the 30 assets are specific to Wales with another (the ONS) having information on Wales. Likewise, the assets come in an accessible form (26 out of 30) are structured, making for easier access to the data within. Accordingly, there is ample opportunity to gain further insight into Wales through the data that is already available. Moreover, if data from private sources can be uncovered, then there could be an even greater wealth of data on Wales.

Data Skills Findings

SAPIO RESEARCH

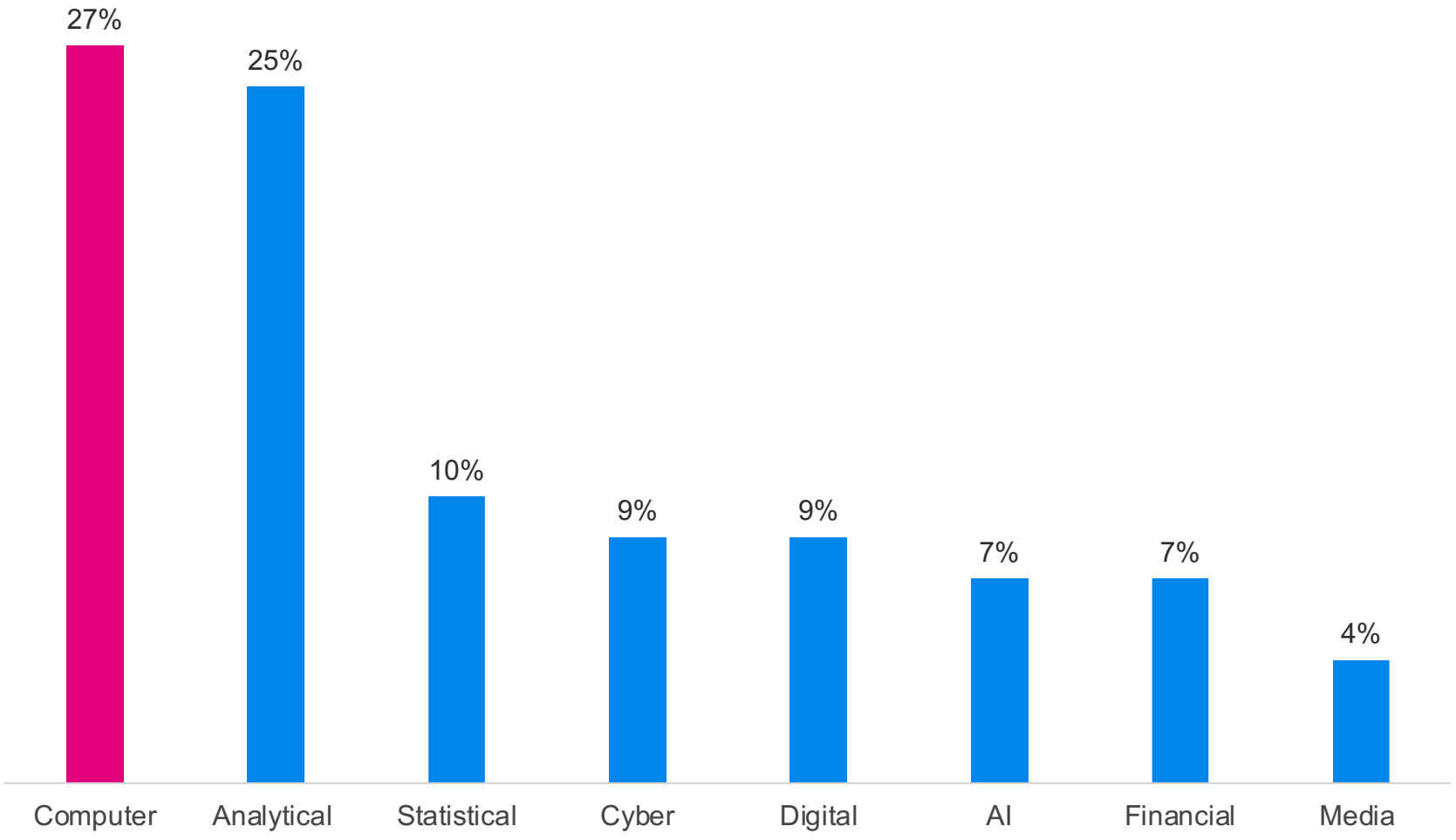
39% of the courses that teach data skills come under the Creative and Professional Services umbrella, whist 37% relate to skills for Future Manufacturing and Systems



Base: 67

N.B The numbers provided here have been calculated using the base number of courses (67) and filtering the entries to provide more granular breakdowns.

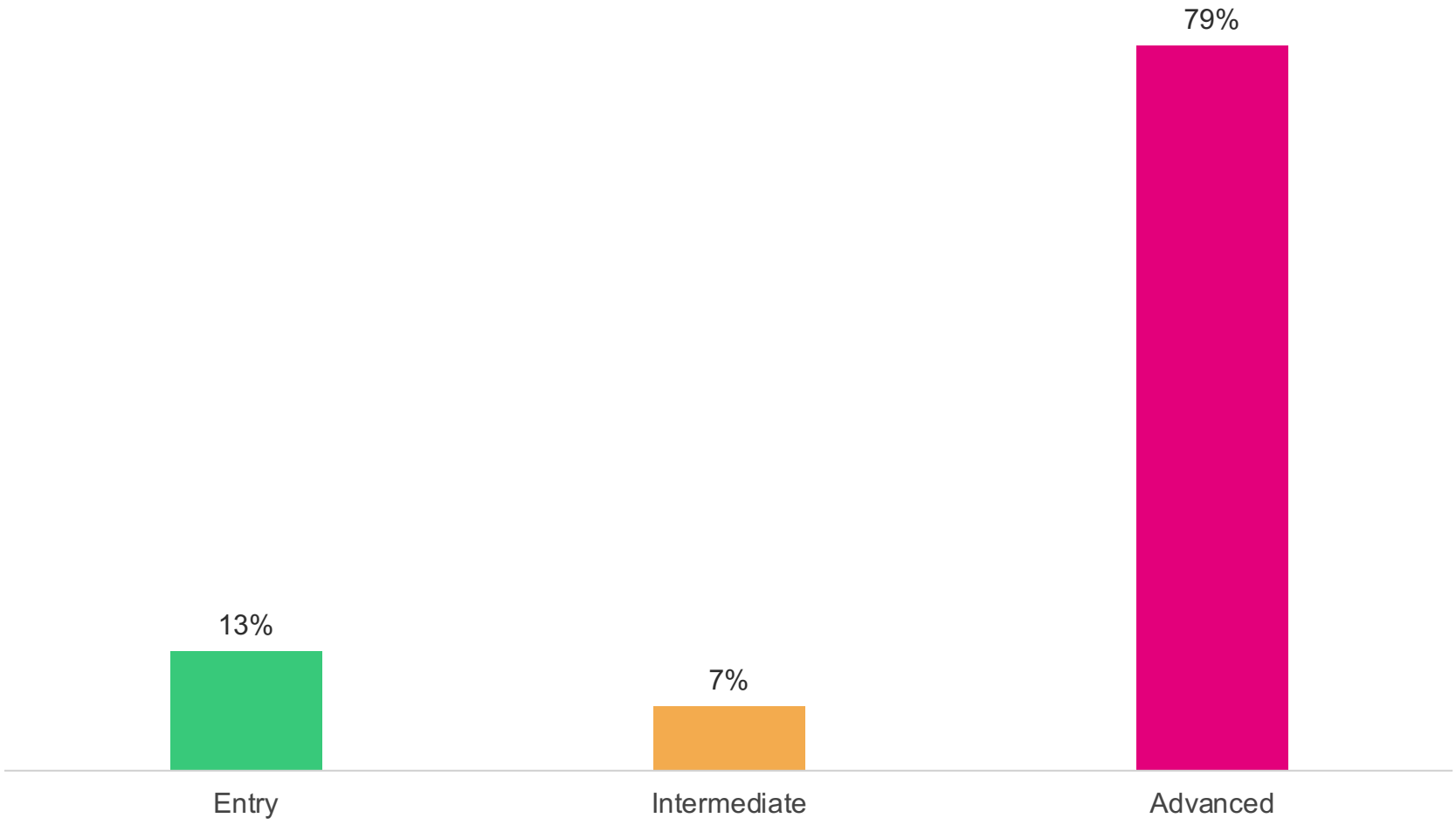
27% of the courses that develop data skills would be considered as helping to improve Computer literacy



Base: 67

N.B The numbers provided here have been calculated using the base number of courses (67) and filtering the entries to provide more granular breakdowns.
** See appendix 2 for definitions of terms*

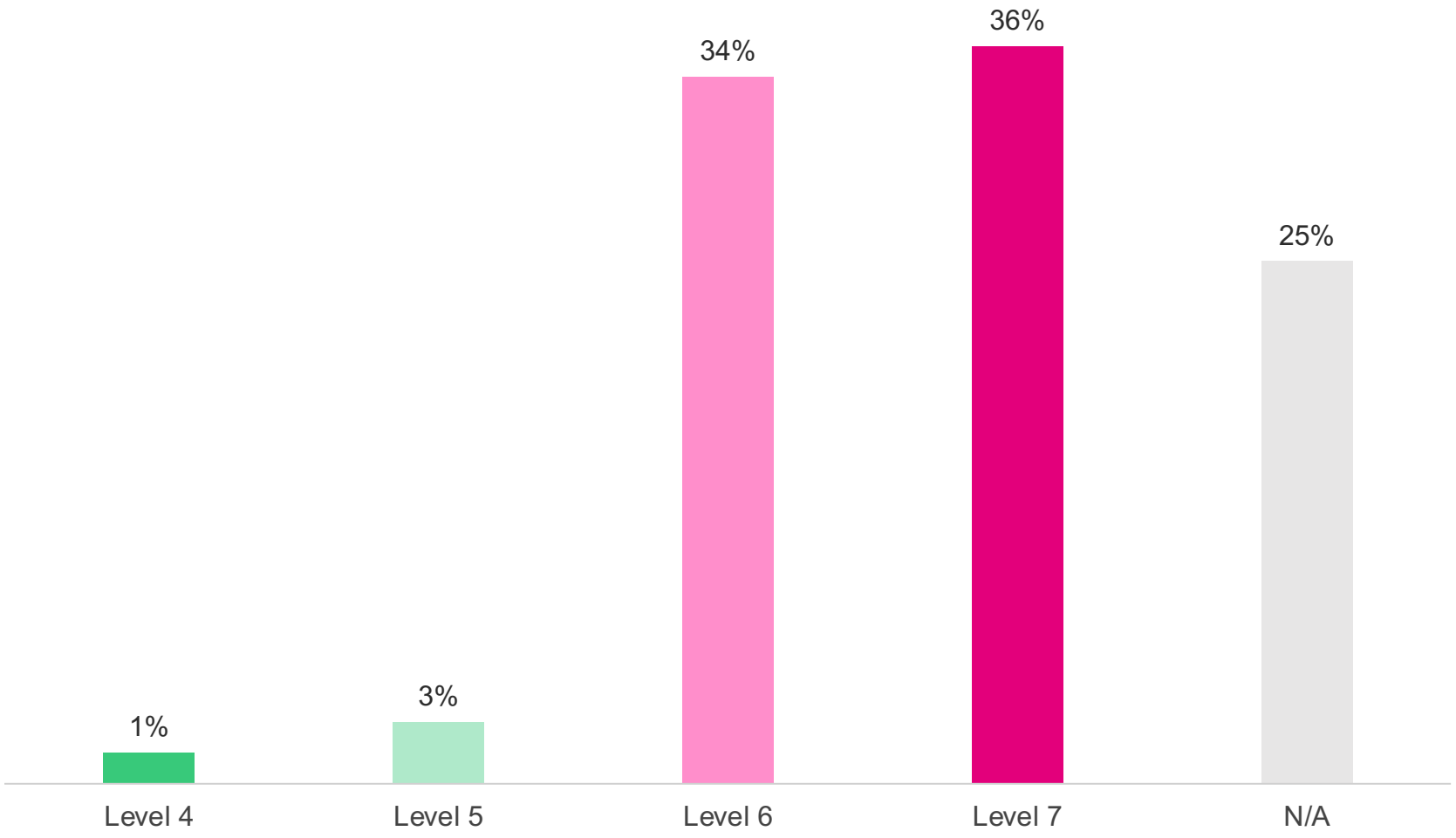
More than three quarters of the courses (79%) are advanced, meaning that they are undergraduate degrees or higher



Base: 67

N.B The numbers provided here have been calculated using the base number of courses (67) and filtering the entries to provide more granular breakdowns.

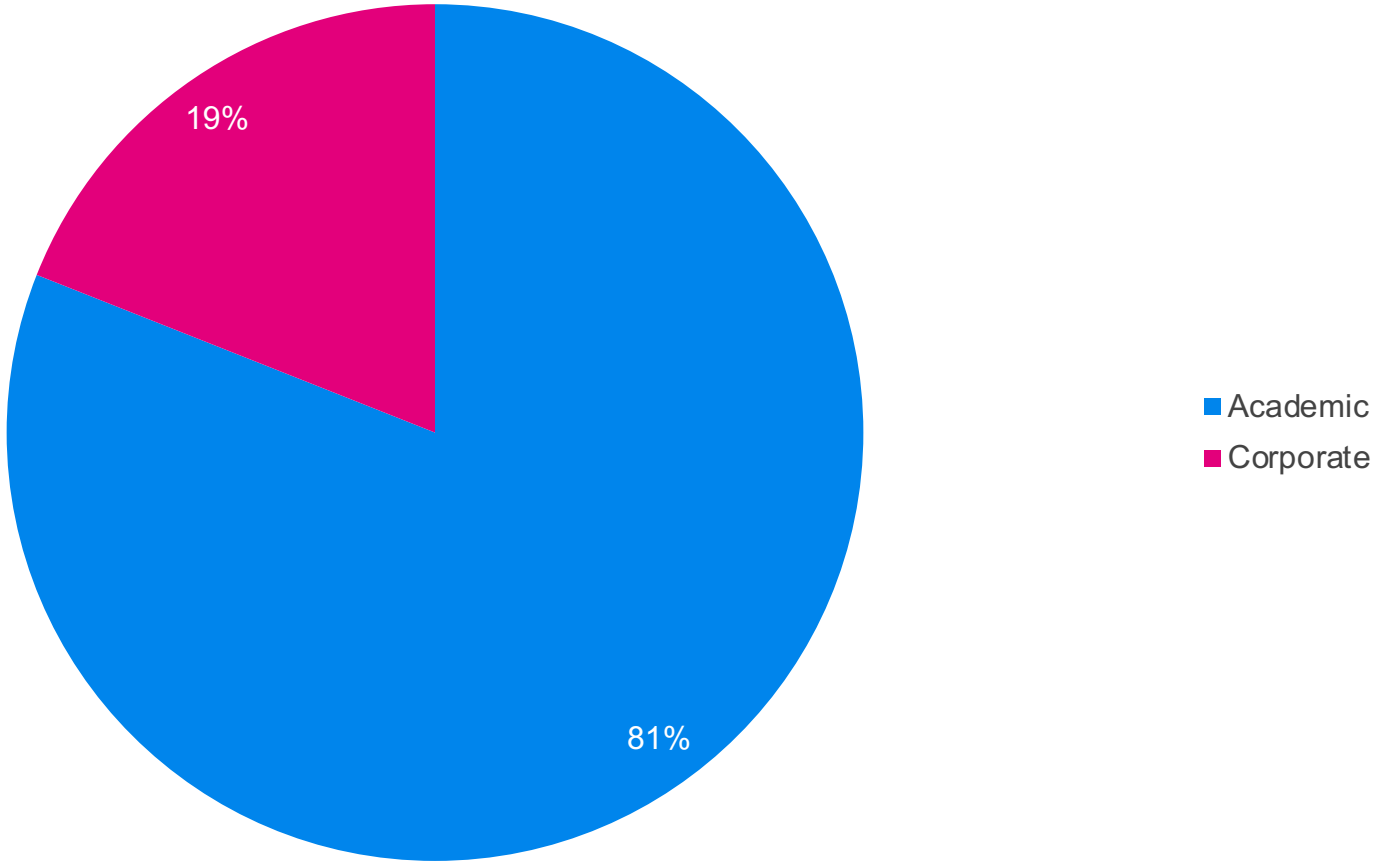
Upon successful completion, 50 out of the 67 courses offer a qualification according to government specifications, with 36% of those being at Level 7 (Masters degree or equivalent)



Base: 67

N.B The numbers provided here have been calculated using the base number of courses (67) and filtering the entries to provide more granular breakdowns.

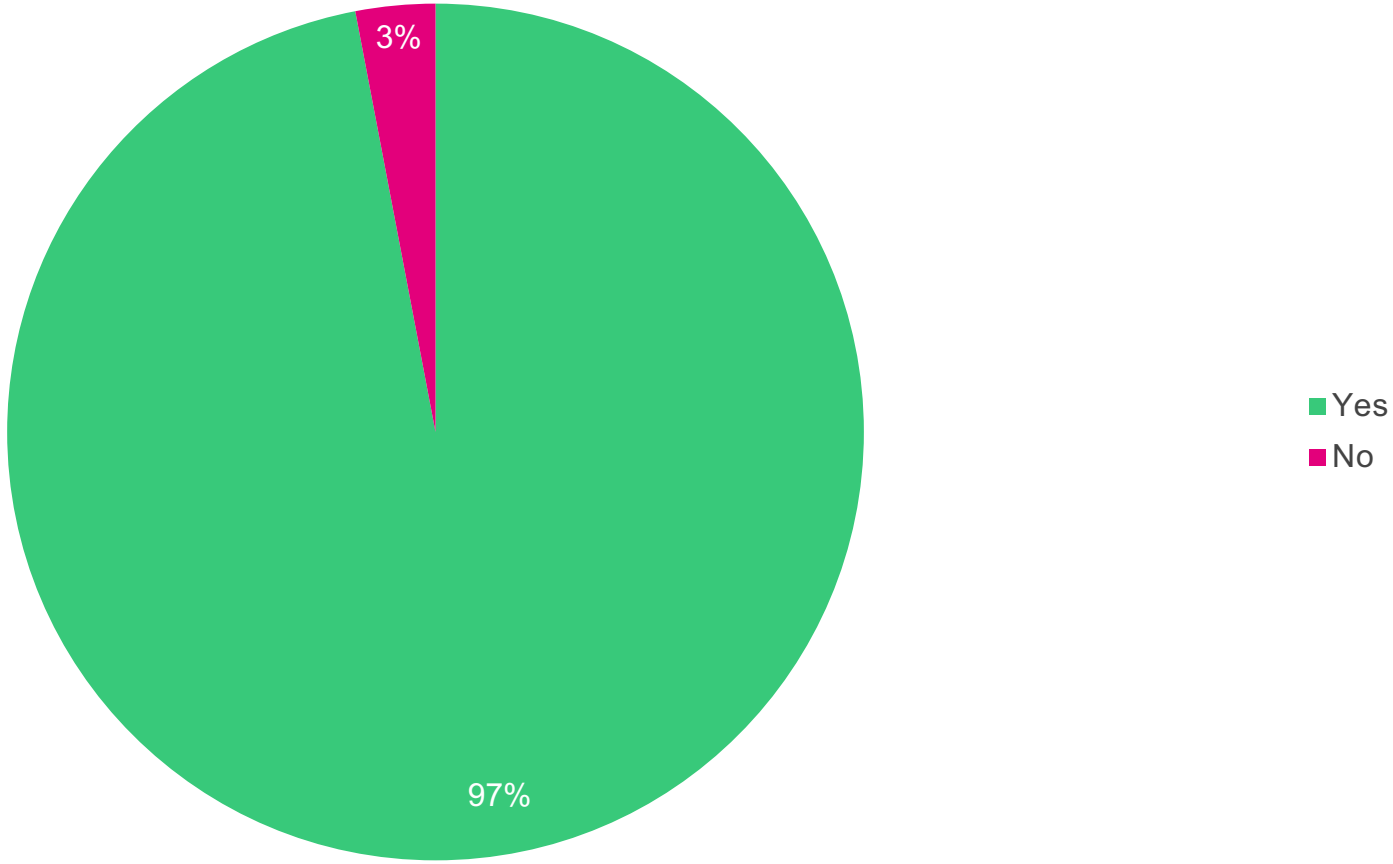
Fewer than 20% of courses are from corporate providers (i.e., non-academic institutions)



Base: 67

N.B The numbers provided here have been calculated using the base number of courses (67) and filtering the entries to provide more granular breakdowns.

More than 95% of the courses identified are available in Wales



Base: 67

N.B The numbers provided here have been calculated using the base number of courses (67) and filtering the entries to provide more granular breakdowns.

Summary

- **Data skills are mainly at an advanced level:** There is an imbalance for the level of data skills in Wales. 53 out of 67 data skills courses were listed as advance (BA/BSc or higher). According to government specifications, 70% of courses are Level 6 or 7. This means that data skills are not being developed at an entry level.

- **Academic institutions dominate training:** The imbalance in data skills continues with who is providing the training course. Academic institutions dominate (54 out of 67) the provision of these courses, again signaling that the opportunities for acquiring basic and intermediate data skills are limited.

- **Wales does develop data skills:** On the plus side, 97% of the courses that develop data skills are available in Wales. Therefore, along with the 67 courses identified, there are opportunities to develop data skills, but predominantly at an advanced level.

- **The courses currently available target different skillsets:** The courses are well distributed in terms of the different types of literacy they come under. That said, courses that cater to analytical (25%) and computer literacy (27%) are most available.

- **Not all skills mirror the WDNA themes:** Data skills do not relate to all the themes of the WDNA. 37% catered to Future Manufacturing and Systems and 39% to Creative and Professional Services. Meanwhile, only 1% is for Public Services and Innovation whilst Net-zero and the Environment is not represented.

Case Studies

Eoin Bailey – UK Innovation Manager at Celsa Steel UK
Key Takeaways

Data Assets

- **Absence of Data is not an issue:** There is a great deal of data being collected but the issue is that it is siloed, even within organisations. Different departments sometimes struggle to access each other's data. Accordingly, the disconnect between departments complicates the process of effective and efficient data use.
- **Using data requires collaboration:** With data currently siloed, it is not always possible to use data correctly. Therefore, a collaborative approach is essential to ensure that the data collected is used to its full potential. In doing so, different skillsets can be brought together to harness and interpret data in different ways

Data Skills

- **Different jobs require different data skills:** Siloed data has led to siloed data skills. People are equipped to deal with the data involved in their job but, outside of that, people are not necessarily well-equipped to handle data. In fine, the different skillsets do not communicate well with each other.
- **Clear communication is essential**: Anyone in Wales who is using data has a vested interest in maximising it for Wales's advantage. To do this, however, there needs to simplicity in communication so that everyone is on the same page when it comes to using data. This translates to simplifying the rhetoric around data so that the nuances of data skills become more accessible to all.

Eoin Bailey – UK Innovation Manager at Celsa Steel UK

Noteworthy Statements

"We certainly don't analyze all the data we may capture effectively"

[We are] "constantly questioning where we can improve, where we can utilize what we have, but also then improve upon the systems in which we're capturing that information."

"There's an opportunity to connect those different silos to actually open them up a little bit and actually look at the, the whole collaborative systems approach."

Dr Marlen Komorowski– Impact Analyst, Cardiff University School of Journalism
Key Takeaways

Data Assets

- **Barriers to accessing data:** Specifically for the creative industries, there is a significant amount of data that is siloed and is therefore inaccessible. With the creative industries increasingly keen to work with data, its inaccessibility is hamstringing their progress.
- **Mix of data formats:** Of the data that can be accessed, another problem encountered is large amounts of unstructured data. Whilst there is data available that comes in accessible form, some require significant cleaning before it can be used, further complicating the process of using data.

Data Skills

- **An interdisciplinary approach is advantageous:** The creative industry contains characters from disparate background and thus have differing data skillsets. In order to maximise the data, different data skillsets should come together to create a collaborative environment for data analysis.
- **Cater to the inexperienced:** There is no shortage of people with the skills to conduct advanced data analysis. Rather, many lack the more basic skills that will allow them to perform simple analysis. Indeed, for those without basic data skills, there is a fear of data and so this stigma needs tackling so that more people have data skills.

Dr Marlen Komorowski– Impact Analyst, Cardiff University School of Journalism

Noteworthy Statements

"It's not about teaching data skills, it's more about enabling them to have just the basic idea and being able to work with people who have the right skills to build something with the data together."

"One of the problems is because it's very hard to find data and measure the sector, then it's also hard to advocate for the sector"

"It's very hard to access the right data about and for the creative industries due to all of these silos"

"Bringing all of these different data skills together on the one hand and then having people with without data skills being able to work with them. That's what makes the difference, in my opinion."

Dr Marlen Komorowski– Impact Analyst, Cardiff University School of Journalism

Creative Economy Atlas Cymru

- The Creative Economy Atlas Cymru is a resource that is showcasing the creative might of Wales

- The Atlas highlights Wales' "strengths and explore its depths, to help make connections and to be a useful tool for those who want to learn more about Wales' creative ecosystem"

- It serves as an excellent example of how data pulled from disparate sources can be made accessible

- The Atlas is also an excellent resource for finding additional data sources, particularly on the creative industry

- It is also entirely focused on Wales, thus providing further granular detail on Wales as a nation

Interviews

Anonymous C-Suite Executive based in South Wales working in Fintech
Key Takeaways

Data Assets

- **There can always be more done data:** Having data is one thing, but maximising it is another. For private companies collecting their own data, there is always scope to get more out of it. There is, therefore, this struggle between having data and being able to make the most of it. This can to some extent be attributed to not have the correct data skills but also lacking the required personnel
- **Private companies have proprietary data:** The search for data will encounter an issue when it come to private companies, namely that they simply cannot release their data. In such situations, it is because the data contains sensitive information that has been entrusted to that company. Therefore, it simply cannot be made public.

Data Skills

- **Keeping existing talent is a problem:** With the advanced data skills courses available, talent is being created and nurtured in Wales. However, there are not enough positions that utilise these skills to encourage the talent to stay in Wales. They are, therefore, leaving in search of opportunities. Likewise, when trying to fill roles where specific skillsets are needed, companies are having to look outside Wales since there simply are not the people available in Wales
- **Internal training uses external resources:** Private companies want to develop and expand the skillsets of their employees. To do this, they will lean on external expertise to form the structure of the training. However, given that this training conducted internally, it is more challenging to learn exactly what skills are being developed as the course are only available to people within that company.

Anonymous C-Suite Executive based in South Wales working in Fintech

Noteworthy Statements

"The feedback we've been getting from recruiters is that candidates for that (role) have been quite difficult to find"

"There is certainly a lot of scope still you know, for particularly practical applied AI activities to be to be grown. There's lots of centers of research where people do extremely abstract things. But you know how those turn into real understandable applied concepts is is very important."

"We don't collect a whole lot to be honest, we work with a lot but most of the data we work with is proprietary data that's owned by our customers.

Anonymous Academic working in Healthcare
Key Takeaways

SAPIO RESEARCH

Data Assets

- **Anonymised data is unavoidable:** When working with certain data assets, anonymity is unavoidable. In fact, it is essential. This is because the data contains sensitive information that simply cannot be shared. Similarly, access is restricted, coming from academic institutions rather than public sources. This means that having full data accessibility cannot be possible.
- **Impact lies in international data:** Undoubtedly data on Wales can have an impact. Yet, Welsh data on its own should not be expected to have the same impact as data collected from around the world. The larger, international studies are more influential. That said, Welsh data could potentially be part of these data and therefore contribute to this impact.

Data Skills

- **Lack of specialisation:** Wales does not lack for courses that train students in advanced data skills, however there is not always the same degree of specialization. The courses could be considered relatively general, albeit advanced. What this means is that there are not the provisions for students in Wales to acquire the skills to work with incredibly nuanced data. This in turn leads to a talent drain with students forced to go elsewhere to acquire these skills.
- **More people needed:** Tying into the absence of courses to train students in specialised data skills is a lack of people to fill these advanced and specialised roles. With ever growing quantities of data that needs to be used, the need for people with the required skills is growing, and it is only becoming more pressing. To that end, attempts need to be made to train people in Wales so that they can then fill the roles available in Wales.

Anonymous Academic working in Healthcare
Noteworthy Statements

SAPIO RESEARCH

"We need to develop more people with those skills because it's only going to become more and more pressing"

"On an international scale, it's sort of feeds into larger efforts and then those larger international studies have much greater power"

"We do have one specific one (course) here because we recognize that that is an area that we do need to develop"

Executive Summary

The State of Data Assets and Data Skills in Wales

Data Assets:

- Data on private companies is not easily accessible both because it is siloed, and it is not advertised that it is available
- Data on Future Manufacturing and Systems is not readily available from public sources, and it is unclear where it might be
- There is data available that specifically relates to Wales and, from the assets found, it also comes in structured form
- Some data contains PPI and is cannot be made public
- Overall, there are already data assets available on Wales and further exploration into the private sector could uncover more
- The main issue to combat is siloed data that cannot be easily accessed

Data Skills:

- Wales excels in catering to and developing advanced data skills at undergraduate level and beyond
- This is true for that skills for Analytical and Computer literacy
- Where the problem lies is in the lack of provision of entry/beginner courses in data skills
- Likewise, there are limited opportunities to develop data skills outside of academic institutions
- That said, it should not be ignored that Wales does offer a considerable number of courses for data skills, just predominantly at an advanced level

Data Assets: SWOT Analysis

Strengths:

- There are considerable assets that provide data on the public sector in Wales
- Much of the data is open-source, allowing for easy access to a lot of data on Wales
- Of the sources found, the majority have structured data, meaning that little cleaning is needed to utilise it
- Many of the assets have data on health, thus providing ample opportunity into the health of Wales as a nation

Weaknesses:

- Data on future manufacturing is difficult to obtain, complicating insight into this area
- Data for the private sector is limited, due to siloing and confidentiality, thus limiting the access to certain data sets
- The uneven distribution of data between themes means that it is hard to get a comprehensive picture of the state of data assets in Wales
- The data from private companies can be proprietary, meaning it cannot be made public under any circumstances

Opportunities:

- Private companies use external resources to conduct data training. If these resources can be learnt, then a clearer picture of data skills courses can be uncovered
- Given the amount of health data available, there is the opportunity to do a lot with this data
- The amount of public and open-source data means that much can be done with it whilst attempting to get access to private or paid data
- Encouraging a collaborative approach to using data would help to maximise its potential

Threats:

- Failure to build develop relationships to obtain private data will lead to an incomplete picture of data and Wales
- Likewise, not looking for more data on future manufacturing and systems would exacerbate the current imbalance of data assets
- Given that much of the data is siloed, failure to address this would potentially undermine any progress made on locating and using data assets
- Ignoring data that comes from outside of Wales and relying entirely on Welsh-specific data could limit the impact of the work

Data Skills: SWOT Analysis

SAPIO RESEARCH

Strengths:

- The number of advanced level courses that teach data skills is remarkable. The mechanisms are in place to teach data skills at an advanced level.
- The courses available cater especially to creative and professional services, and future manufacturing and systems, thus creating a collection of individuals that can maximise the data available in these areas
- Almost all the courses identified are in Wales, indicating that there is already the infrastructure in Wales to develop data skills
- Skills are being developed for computer and analytical especially, suggesting that Wales has a particular strength for data skills

Weaknesses:

- Skills are predominantly being developed at an advanced level and less is being done to train people in more basic data skills
- Courses to develop these skills take place at universities, further limiting the opportunities for more people to develop data skills
- Although data skills are developed in Wales, the people with them often leave due to a lack of opportunities in Wales
- The language around data skills can be highly nuanced and this can create a certain fear factor around data, thus discouraging people from seeking further training

Opportunities:

- Although finding private data could be more challenging, getting access to it could provide invaluable insights
- By either finding or starting courses that develop basic data skills, more people can become data literate
- More people with data skills are always needed so by advertising the bounty of data skills courses, more data literate people will be developed in Wales
- Specialist data skills at an advanced level are needed so there is the opportunity to use the infrastructure already in place at universities to develop courses that can do this

Threats:

- Perhaps the biggest threat is the talent drain. Failure to address this will mean that the results of the advanced data courses are being enjoyed outside of Wales
- Without catering to the inexperienced, the imbalance between basic and advanced data skills will remain
- Maximising data requires an interdisciplinary approach. At present skills are somewhat siloed in their own niches which in turn creates a lack of harmony and cohesion between skillsets
- At present, there can be a fear factor around data with people concerned that they are not skilled enough to use it. Without addressing this fear factor, developing data skills among more people will remain challenging

Recommendations

Data Assets:

- Attempt to forge relationships with private companies that can provide data for the key themes of the WDNA
- Target organisations that can provide data in the areas where there are currently a lack of assets
- Attempt to find innovative solutions to resolve the problem caused by data assets containing PPI
- Try to understand why data is being siloed and consider ways of overcoming this hurdle

Data Skills:

- Understand the need for more entry level courses that target a wide range of data skills, not just data analysis, data science etc.
- Consider possible methods for providing data skills to more people
- Develop plans that take into account differing data skills and the importance of collaboration

Thank you!

Sapio Research
Pentagon House,
52-54 Southwark Street,
London, SE1 1UN

t: (+44) 0207 2361 604
e: team@sapioresearch.com
www.sapioresearch.com

Audience | Brand | Content Research

Appendix 1

- ***Health and Wellbeing:*** Precision medicine, diagnostics and interventions, intelligent healthcare systems, social care through AI
- ***Net-zero and the Environment:*** Energy and transport, environmental management, circular and green economies, housing, agri-tech
- ***Public Services Innovation:*** Intelligence, efficiency, automation, enhanced decision making, advanced problem solving, personalisation
- ***Future Manufacturing and Systems:*** Factory of the future, advanced materials, resilience in supply chains, digital twins, smart manufacturing, agri-tech
- ***Creative and Professional Services:*** Legal, financial technology, business systems, social media, human centred systems and communication

Appendix 2

- ***Financial Literacy:*** knowledge needed to make important financial decisions
- ***Digital Literacy:*** skills to understand how to use the internet
- ***Computer Literacy:*** the knowledge and ability to use computers and related technology efficiently, with skill levels ranging from elementary use to computer programming and advanced problem solving
- ***Media Literacy:*** includes critical thinking, evaluation strategies, knowledge of operation of news and media industries, along with defining digital media literacy as being able to distinguish fact from fiction, including misinformation, understand how digital platforms work, as well as how to exercise one's voice and influence decision makers in a digital context
- ***AI Literacy:*** competencies necessary in a future in which AI transforms the way that we communicate, work, and live with each other and with machines
- ***Statistical Literacy:*** the skills to make decisions based on numerical data that confronts us in all aspects of our personal and professional lives
- ***Cyber Literacy:*** knowing where to go to find reliable and accurate resources on the internet
- ***Analytical/Evaluation Literacy:*** having the skills to understand evidence before making decisions