# Methodological notes for start-ups: entry, early-stage funding and successful exit

International Energy Agency

# General considerations about Crunchbase and tagging of sectors or business strategies

The source of data for the database on start-ups is Crunchbase. Crunchbase provides a depth and breadth of knowledge that ensures its data is recognised as the primary source of business information by over 55 million users globally.

The data are sourced through two main channels: a large investor network and community contributors. In addition, executives, entrepreneurs, and investors contribute to update and revise Crunchbase company profile pages. The data is processed with artificial intelligence and machine learning algorithms in order to ensure accuracy and scan for anomalies. Additionally, algorithms continuously search the web and thousands of news publications for information to enrich profiles.

The crowd-sourcing process, the partnership with investment firms, and the validation with artificial intelligence and machine-learning algorithms represent important innovations compared to other commercial databases and public data sources commonly used in economic research, which may provide unprecedented opportunities to analyse phenomena that have been under investigated so far because of lack of suitable data.

As of 2007 Cruncbase provides a database of 1.6 million start-ups with many associated tables that include useful elements such as:

- Founding dates of start-ups
- Funding amounts and types throughout their lifetime
- Description of the start-ups, including text and sector classification
- Details of the IPO (when relevant)
- Details of the acquisition (when relevant)
- People working in the company

To describe start-ups, Crunchbase also provides a labelling of more than 700 general “tags” that are provided by the firms and/or Crunchbase staff. It is





This report is part of the IEA’s Clean Energy Transitions in Emerging Economies programme, which has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 952363.

important to note that individual start-ups can have multiple tags, with a mean of 2.9 tags per start-up. 24 of these tags are clearly energy-related in nature, and these are listed below.

**Table 1. Crunchbase categories related to energy**

| Crunchbase categories |
|---|
| Battery |
| Biofuel |
| Solar Energy |
| Clean Energy |
| Electric Vehicle |
| Electrical Distribution |
| Energy |
| Energy Efficiency |
| Energy Management |
| Energy Storage |
| Fossil Fuels |
| Fuel |
| Fuel Cell |
| Green Building |
| Lighting |
| Nuclear |
| Oil and Gas |
| Power Grid |
| Renewable Energy |
| Smart Building |
| Smart Cities |
| Smart Home |

In order to further elaborate those fields, we used a text recognition algorithm to identify keywords used in the description text of the start-up provided by the firms themselves. Those sectors were then aggregated into the more harmonised and aggregated energy sectors below:





This report is part of the IEA's Clean Energy Transitions in Emerging Economies programme, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 952363.

**Table 2. Aggregated energy categories**

| Crunchbase categories |
|---|
| Storage (not electromobility) |
| Bioenergy |
| Solar |
| Other renewables |
| Hybrid and Electric mobility |
| Grid |
| Energy Efficiency |
| Fossil Fuels |
| Hydrogen and Fuel Cells |
| Nuclear |
| Wind |

We also allocated firms with other Crunchbase tags into other aggregates[1] that are likely to be complementary, and are likely to be of policy interest, namely:

- Digital: all start-ups that are related to digitalisation. It includes topis like Internet of Things, IT systems, software, applications, social media etc…
- Green: All start-ups that are working towards sustainable development. This includes topics such as pollution control, waste management, greens consumer goods but also topics that are related to the energy sectors such as renewable energy, green buildings, electric vehicles etc.

The objective of this is to analyse the penetration of these orthogonal business strategies within energy and non-energy start-ups.

[1] Further information on the description of these are available from the authors upon request.





This report is part of the IEA's Clean Energy Transitions in Emerging Economies programme, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 952363.

# Indicators on start-up creations

The first set of indicators is related to start-up creations. Those indicators will give a measure of the amount of entrepreneurship within the relevant sector (as identified by the tags). Scripts were written to enable the generation of four sets of indicators:

1. Start-ups in energy: Energy and non-energy start-up creation per country and year in both numbers and shares
2. Start-ups in energy and digital: Energy and digital start-up creation per country and year in both numbers and shares
3. Start-ups in energy and green: Energy and green start-up creation per country and year in both numbers and shares
4. Start-ups in energy technologies: Energy start-up creation per country, year and sub-energy sector in both numbers and shares

The scripts that generate the first three datasets are written following this logic:

1. Automated text-analysis of all the start-up descriptions, in order to elaborate Crunchbase sectoral classification (740 tags).
2. Identification of the presence of the high-level sectors energy, green and digital with dummy variables for all start-ups
3. Aggregation by year, country and high-level sector dummies
4. Aggregation for the four relevant sector combinations (energy, non-energy, energy and green, energy and digital)
5. Generation of the indicators
6. The script for the fourth dataset is generated following this logic:
7. Automated text-analysis of all the start-up descriptions, in order to elaborate Crunchbase sectoral classification (740 tags).
8. Identification of energy-related sectors
9. Ventilation of start-ups (see note below) into fractional counts (based on presences of different tags) corresponding to their respective energy sectors.
10. Aggregation by year, country and energy field using all the fractional counts
11. Generation of the indicators





This report is part of the IEA's Clean Energy Transitions in Emerging Economies programme, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 952363.

**Note on ventilating data for start-up**

To generate fractional counts of the energy sectors within each company, we used the ventilation method. This consists in exploding each company into as many pieces as their number of energy sectors. All pieces will be allocated a weight corresponding to the number of occurrences of the given sector. The weights will be normalised so that the sum of the weights is equal to 1. Those weights will be called "fractional counts". Note that a different fractional count will be used for patents.

**Table 3. Metadata for start-ups counts at a high level or sub-energy level**

| Column | Explanation |
|---|---|
| year | Founding year of the start-ups. |
| country_ISO | Country of the start-up in ISO3 format |
| Indicator categories | Name of the indicator used |
| Technology or Sector | High-level sector used to group start-ups (including non-energy, energy-and-digital, energy-and-non-digital, energy-and-green, energy-and-not-green) or energy technology field (wind, solar, etc…) |
| Value | Number of companies in the sector, country and year. |

The need for 'nowcasting' is particularly important in the case of start-up creation data. When start-ups are created it is hard to identify them in a timely manner – they are too small to be captured in many traditional data sources and often fall through the cracks of data identification. This means that they are added in databases with a certain time lag. In many cases they disappear even before they are identified. Machine learning methods were applied to "predict" the lagged incorporation of unobserved start-ups in the data set. Given the need for a





This report is part of the IEA's Clean Energy Transitions in Emerging Economies programme, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 952363.

reasonably large sample size this could only be done at the level of broad aggregations[2].

# Indicators on investment in start-ups

This set of indicators focuses on the level of investment allocated to start-ups in the energy sector and beyond. It will allow the evaluation of the share of investment in energy start-ups in total start-up investment as well as the level of start-up investment observed in the different Energy fields. The indicators focus on entrepreneurial firms – also called *de novo* – thereby excluding spin-offs from incumbent firms. The scripts create the following datasets:

- Fundings in energy: investment in energy and non-energy start-ups per country and year in USD 2020 and share (%)
- Fundings in energy-digital: investment in energy-digital start-ups per country and year in USD 2020 and share (%)
- Fundings in energy-green: investment in energy-green start-ups per country and year in USD 2020 and share (%)
- Fundings by technology: investment in energy fields per country and year in USD 2020 and share (%)
- Likelihood to receive funding: probability for a start-up in a given country and sector to receive different types of funding (early stage, expansion stage and bridge stage)
- Time to receive funding: average time in years for a start-up in a given country and sector to receive different types of fundings (early stage, expansion stage and bridge stage)

To generate the datasets, we used Crunchbase's funding rounds table and simplified the funding rounds classification using the following mapping table:

**Table 4. Investment types in Crunchbase and their corresponding investment stage**

| investment_type | investment_stage |
|---|---|
| private_equity | Bridge stage |
| corporate_round | Bridge stage |

[2]For further information about how start-up creation was "nowcasted" see report "Generating timely energy market and behavioural data with innovative methods".





This report is part of the IEA's Clean Energy Transitions in Emerging Economies programme, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 952363.

| investment_type | investment_stage |
|---|---|
| grant | Bridge stage |
| debt_financing | Bridge stage |
| post_ipo_equity | Bridge stage |
| post_ipo_debt | Bridge stage |
| post_ipo_secondary | Bridge stage |
| equity_crowdfunding | Crowdfunding |
| product_crowdfunding | Crowdfunding |
| series_a | Early stage |
| seed | Early stage |
| series_b | Early stage |
| pre_seed | Early stage |
| angel | Early stage |
| convertible_note | Early stage |
| initial_coin_offering | Early stage |
| series_unknown | Expansion stage |
| secondary_market | Expansion stage |
| series_c | Expansion stage |
| series_d | Expansion stage |
| series_e | Expansion stage |
| series_f | Expansion stage |
| series_g | Expansion stage |
| series_h | Expansion stage |
| series_i | Expansion stage |
| series_j | Expansion stage |
| undisclosed | Unknown |
| non_equity_assistance | Early stage |

**Definitions of investment types**

**Angel**: An angel round is typically a small round designed to get a new company off the ground. Investors in an angel round include individual angel investors, angel investor groups, friends, and family.

**Pre-Seed**: A Pre-Seed round is a pre-institutional seed round that either has no institutional investors or is a very low amount, often below USD 150 000.





This report is part of the IEA's Clean Energy Transitions in Emerging Economies programme, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 952363.

**Seed**: Seed rounds are among the first rounds of funding a company will receive, generally while the company is young and working to gain traction. Round sizes range between USD 10 000 k to USD 2 million, though larger seed rounds have become more common in recent years. A seed round typically comes after an angel round (if applicable) and before a company's Series A round.

**Venture - Series Unknown**: Venture funding refers to an investment that comes from a venture capital firm and describes Series A, Series B, and later rounds. This funding type is used for any funding round that is clearly a venture round but where the series has not been specified.

**Series A and Series B** rounds are funding rounds for earlier stage companies and range on average between USD 1 million to USD 30 million.

**Series C, D, E, F, G, H, I** rounds and onwards are for later stage and more established companies. These rounds are usually USD 10 million+ and are often much larger.

**Equity Crowdfunding**: Equity crowdfunding platforms allow individual users to invest in companies in exchange for equity. Typically, on these platforms the investors invest small amounts of money, though syndicates are formed to allow an individual to take a lead on evaluating an investment and pooling funding from a group of individual investors.

**Product Crowdfunding**: In a product crowdfunding round, a company will provide its product, which is often still in development, in exchange for capital. This kind of round is also typically completed on a funding platform.

**Private Equity**: A private equity round is led by a private equity firm or a hedge fund and is a late stage round. It is a less risky investment because the company is more firmly established, and the rounds are typically upwards of USD 50 million.

**Convertible Note**: A convertible note is an 'in-between' round funding to help companies hold over until they want to raise their next round of funding. When they raise the next round, this note 'converts' with a discount at the price of the new round. You will typically see convertible notes after a company raises, for example, a Series A round but does not yet want to raise a Series B round.

**Debt Financing**: In a debt round, an investor lends money to a company, and the company promises to repay the debt with added interest.

**Secondary Market**: A secondary market transaction is a fundraising event in which one investor purchases shares of stock in a company from other, existing





This report is part of the IEA's Clean Energy Transitions in Emerging Economies programme, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 952363.

shareholders rather than from the company directly. These transactions often occur when a private company becomes highly valuable and early stage investors or employees want to earn a profit on their investment, and these transactions are rarely announced or publicised.

**Grant**: A grant is when a company, investor, or government agency provides capital to a company without taking an equity stake in the company.

**Corporate Round**: A corporate round occurs when a company, rather than a venture capital firm, makes an investment in another company. These are often, though not necessarily, done for the purpose of forming a strategic partnership.

**Initial coin offering (ICO)**: An initial coin offering (ICO) is a means of raising money via crowdfunding using cryptocurrency as capital. A company raising money through an ICO holds a fundraising campaign, and during this campaign, backers will purchase a percentage of a new cryptocurrency (called a "token" or "coin"), often using another cryptocurrency like bitcoin to make the purchase, in the hopes that the new cryptocurrency grows in value.

**Post-IPO Equity**: A post-IPO equity round takes place when firms invest in a company after the company has already gone public.

**Post-IPO Debt**: A post-IPO debt round takes place when firms loan a company money after the company has already gone public. Similar to debt financing, a company will promise to repay the principal as well as added interest on the debt.

**Post-IPO Secondary**: A post-IPO secondary round takes place when an investor purchases shares of stock in a company from other, existing shareholders rather than from the company directly, and it occurs after the company has already gone public.

**Non-Equity Assistance**: A non-equity assistance round occurs when a company or investor provides office space or mentorship and does not get equity in return.

**Funding Round**: "Funding round" is the general term used for a round when information regarding a more specific designation of the funding type is unavailable.

We converted those funding round type into the following broad stages in order to favour aggregation and analysis.





This report is part of the IEA's Clean Energy Transitions in Emerging Economies programme, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 952363.

**Early-stage:** This stage encompasses all investments from the start-up birth to the start-up market launch. This ends when the company is starting to generate revenues

**Expansion stage funding:** This stage is when the start-up is seeing exponential growth and needs additional funding to keep up with the demands. Because the business likely already has a commercially viable product and is starting to see some profitability, venture capital funding in the emerging stage is largely used to grow the business even further through market expansion and product diversification.

**Bridge stage funding:** In this stage, the start-up has reached maturity. Funding obtained here is typically used to support activities like mergers, acquisitions, or IPOs. The bridge state is essentially a transition to the company being a full-fledged, viable business.

The scripts would generate the datasets with the following logic:

- Identification of *de novo* start-ups
- Collection of the funding rounds from all de *novo* start-ups
- Allocation of the simplified funding types to every funding round
- Allocation of sectors based on the data coming from the indicators on start-up creation (energy, non-energy, energy-green, energy-digital, and the sub-energy sectors)
- Ventilation of all the energy start-ups deal based on the fractional count observed in start-ups[3]
- Aggregation of data per year, sector, country
- Determination of likelihood to receive different types of fundings
- Determination of the time it takes to receive different types of fundings
- Generation of the indicators

**Table 5. Metadata for fundings in start-ups**

| Column | Explanation |
|---|---|
| year | Founding year of the start-ups. |

[3] See part on indicators on start-up creations





This report is part of the IEA's Clean Energy Transitions in Emerging Economies programme, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 952363.

| Column | Explanation |
|---|---|
| country_ISO | Country of the start-up in ISO3 format |
| Indicator categories | Funding stage (early stage, expansion stage, bridge stage) |
| Technology or Sector | High-level sector used to group start-ups (including non-energy, energy-and-digital, energy-and-non-digital, energy-and-green, energy-and-not-green) or energy technology field (wind, solar, etc…) |
| Unit | USD (2020) or % |
| Number of observations | Number of deals observed |
| Value | Amount of money (USD 2020) spent or % share |

**Table 6. Metadata for time to receive fundings**

| Column | Explanation |
|---|---|
| year | Founding year of the start-ups. |
| country_ISO | Country of the start-up in ISO3 format |
| Indicator categories | Funding stage (early stage, expansion stage, bridge stage) |
| Technology or Sector | High-level sector used to group start-ups (including non-energy, energy-and-digital, energy-and-non-digital, energy-and-green, energy-and-not-green) or energy technology field (wind, solar, etc…) |
| Unit | years |
| Value | Average years to receive funding |

**Table 7. Metadata for likelihood to receive fundings**

| Column | Explanation |
|---|---|
| year | Founding year of the start-ups. |
| country_ISO | Country of the start-up in ISO3 format |
| Indicator categories | Funding stage (early stage, expansion stage, bridge stage) |





This report is part of the IEA's Clean Energy Transitions in Emerging Economies programme, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 952363.

| Column | Explanation |
|---|---|
| Technology or Sector | High-level sector used to group start-ups (including non-energy, energy-and-digital, energy-and-non-digital, energy-and-green, energy-and-not-green) or energy technology field (wind, solar, etc…) |
| Unit | % |
| Value | % likelihood to receive funding |

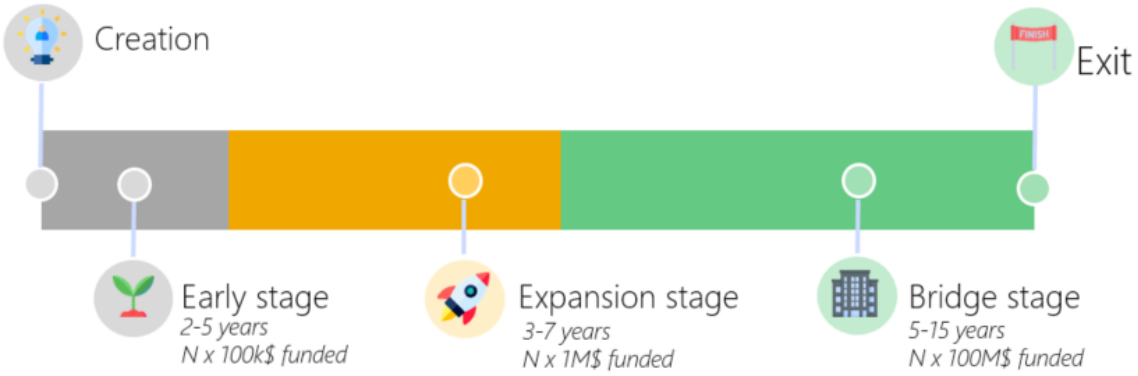
## Indicators on start-up success

This set of indicators will focus on measuring how successful start-ups have been in different countries and sectors. Those indicators will measure the following elements:

- Likelihood to succeed. In this indicator *success* will be considered to be attained when the start-up gets acquired or goes public.
- Time to succeed. This indicator evaluates the amount of time it takes for those start-up that have *succeeded.*

These indicators are useful to track whether firms have experienced a successful exit (IPO or acquisition). Given the nature of the database we cannot measure success in terms of organic growth. As such it is important to bear in mind that firms that have not exited successfully may have experienced organic growth (rather than "failure" per se).

**Figure 1. Theoretical milestones within the lifetime of a start-up. Numbers are indicative**








This report is part of the IEA's Clean Energy Transitions in Emerging Economies programme, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 952363.

The indicators on probability of success have been computed for datasets observed over a 20-year window for the following reasons:

- 20 years is a reasonable time window to allow start-ups to become successful. Based on dataset observations, it is highly unlikely for start-ups to be successful after 20 years.
- A moving time window has been chosen to allow datasets comparison over time. So the indicators are not biased by data volume and give an equal weight to new data that are input into the dataset.

The scripts would generate the datasets using the following logic:

- Retrieve tables IPOs and acquisitions, and merge them with companies.
- Determine success of every company.
- Determine time to IPO and acquisition for each individual company.
- Proceed with aggregations at country, year, field level.
- Compute indicators.
- And for the indicators on probability of success per year reference:
- Loop on 20 years window and recompute the indicators for all windows.

The following datasets are generated:

- Success rates in energy: probability of success for start-ups within 3, 5, 10, 20 years and time to reach success per country and sector (energy, non-energy).
- Success in rates in energy and digital: probability of success for start-ups within 3, 5, 10, 20 years and time to reach success per country and sector (energy and digital, energy and non-digital).
- Success rates in energy and green: probability of success for start-ups within 3, 5, 10, 20 years and time to reach success per country and sector (energy and green, energy and non- green).
- Success rates per technology: probability of success within 3, 5, 10, 20 years for start-ups per country and technology field.
- Probability of success in energy: probability for start-ups to reach milestones[2] per year reference, country and sector (energy, non-energy).
- Probability of success in energy and digital: probability for start-ups to reach milestones[2] per year reference, country and sector (energy and digital, energy and non-digital).





This report is part of the IEA's Clean Energy Transitions in Emerging Economies programme, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 952363.

- Probability of success in energy and green: probability for start-ups to reach milestones[2] per year reference, country and sector (energy and green, energy and non-green).
- Probability of success by technology: probability for start-ups to reach milestones per year reference, country and technology field.
- Time to succeed by technology: average weighted time for start-ups to reach milestones per country and technology field.

**Table 8. Metadata for success rates**

| Column | Explanation |
|---|---|
| year | Founding year of the start-ups. |
| country_ISO | Country of the start-up in ISO3 format |
| Indicator categories | Acquisition, IPO, success |
| Technology or Sector | High-level sector used to group start-ups (including non-energy, energy-and-digital, energy-and-non-digital, energy-and-green, energy-and-not-green) or energy technology field (wind, solar, etc…) |
| Unit | % |
| Value | percentage of success, IPO or acquisition |

**Table 9. Metadata for time to success**

| Column | Explanation |
|---|---|
| year | Founding year of the start-ups. |
| country_ISO | Country of the start-up in ISO3 format |
| Indicator categories | Acquisition, IPO, success |
| Technology or Sector | High-level sector used to group start-ups (including non-energy, energy-and-digital, energy-and-non-digital, energy-and-green, energy-and-not-green) or energy technology field (wind, solar, etc…) |
| Unit | years |
| Value | Time to success, IPO or acquisition |





This report is part of the IEA's Clean Energy Transitions in Emerging Economies programme, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 952363.

**Table 10. Probability of success within n years**

| Column | Explanation |
|---|---|
| year | Founding year of the start-ups. |
| country_ISO | Country of the start-up in ISO3 format |
| Indicator categories | Within 3, 5, 10, 20 years |
| Technology or Sector | High-level sector used to group start-ups (including non-energy, energy-and-digital, energy-and-non-digital, energy-and-green, energy-and-not-green) or energy technology field (wind, solar, etc…) |
| Unit | % |
| Value | Probability to succeed within 3, 5, 10 or 20 years |





This report is part of the IEA's Clean Energy Transitions in Emerging Economies programme, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 952363.