Fathom Documentation

Fathom Contributors

Feb 03, 2019

Contents:

Warning: Fathom is still in the early stages of development. Here there be dragons. It's a great time to get involved though!

Fathom is a cryptoeconomic protocol for creating globally recognized credentials. It is implemented in a collection of smart contracts in Solidity, and deployed to the Ethereum blockchain.

Fathom aims to enable a new open learning ecosystem. On top of it you can create new structures for learning: things like schools, tutoring systems, or peer-to-peer learning networks.

Note: Cryptoeconomics is a relatively new term and field. If you want to go deeper down the rabbit hole, check out awesome-cryptoeconomics from L4.

CHAPTER 1

Why decentralized credentials?

Today most reputable credentials are issued by large institutions. This has led to a credentialing system that is inaccessible to many, and that is slow to adapt to changing realities. In this system, most can only communicate a small part of their global skillset.

With fathom, it is possible to create credentials that capture any kind of knowledge or skill. Moreover, conceiving of and defining credentials is no longer the exclusive domain of centralized entities. Rather, it is something anyone can participate in.

As well, the process of earning credentials is less exclusionary and more meaningful. The fathom protocol defines an assessment-game in which qualified assessors are economically incentivized to come to a truthful evaluatation of an applicant's skill.

As such, any community in any field can create its own credentials and use them to self-organize. Its members are empowered to communicate these skills and others to the outside world.

1.1 Introduction

Society functions on knowing what people can do. Everybody needs to be able to communicate their skills to others in order to coordinate with them. This used to be a social process within a local community, but society has grown and largely outsourced that function to institutions. Today, people need to communicate a greater diversity of skills and experiences than ever before, over larger timescales and across geographic, cultural and linguistic barriers. It is our belief that institutions do not provide that service well, and that they will be increasingly ill-suited to provide it in the future.

We seek to provide an alternative to institutional credentials. Herein is specified a social protocol with economic incentives that enables knowledge communities to define their own standards and individuals to be assessed in those standards, resulting in credentials that are meaningful, verifiable and durable.

1.1.1 Problem Statement

The coupling of learning and assessment in current institutional models is unscalable and creates a set of perverse incentives for both educators and students. The bureaucracy of centralized institutions makes them resource intensive

and slow to adapt to changes. As a consequence, they are only able to offer a small set of experiences which default to those that can be mass-produced.

Because communicating one's experiences is so essential in today's society, it is in an individual's best interest to actively mold their experiences towards what they can communicate instead of what they can aspire to. Therefore, relying on instituions to be the arbiter of people's abilities has a chilling effect on societal progress.

1.1.2 Introducing Fathom

Fathom is a protocol to create and assess meaningful credentials through the consensus of knowledge communities.

It allows anyone to create a credential and anyone to be assessed in it. The core process involves a jury of randomly assembled assessors with relevant experience, as previously proven by the protocol, playing an 'assessment-game' in which they are economically incentivized such that an accurate assessment is the schelling point.

The protocol makes no assumptions about what is being assessed. Instead, it allows communities to form their own definitions and rules, and carry them out collectively.

Implemented on a public blockchain, it will be possible to distribute the work necessary for assessments to scale far beyond what institutions are capable of.

Furthermore, blockchains can enable a credential-ecosystem that is truly inclusive, accessible and extensible; one which is censor-resistant, durable, and that leaves individuals in full control of their identities. Individuals are able to acquire knowledge and accumulate experiences towards their unique aims, while also shaping and strenghtening the network in their role as an assessor.

1.1.3 Overview

The purpose of this document is to provide a formal specification of the fathom-protocol. It is presented in the following parts: First, The Fathom Protocol explains the various elements that make up the fathom network. Then, the process of creating concepts and getting assessed in them is presented step-by-step. Next, Concept Governance and Network Upgrades describes how forking (cloning) is used to collectively steer credential definitions and to issue protocol upgrades. The third part, Security, lays out the structure that incentivizes users to follow the protocol, while making it economically disadvantageous to deviate from it. Also presented here are some potential attack vectors and how they are mitigated. The final part, Vision, ties these threads to the goals and ambitions of the project.

1.1.4 Contribution

We believe in transparency and collaboration. We invite readers to engage with us – through comments, ideas, or by actively taking part in fathom's development. Together, we seek to enable real-world applications of a wide variety to be built on this new layer of trust and digital social relationships.

To learn more about our overall project roadmap, the infrastructure we are building on top of the fathom protocol, our development process, and ways to contribute, please check out our repositories and our blog.

1.2 The Fathom Protocol

1.2.1 Architecture Overview

This paragraph will introduce the three main components of a fathom-assessment: 1. a *concept*, representing an assessable quality 2. the *concept-ontology*, which relates all concepts to each other. 3. *assessments* which draw individuals from the concepts to be assessors and, upon a positive outcome, add new individuals to them.

Concept

Concept is an umbrella term to capture any kind of skill, quality, piece of knowledge or fact that can be established about an individual. Therefore, each concept C has the following properties:

- **Parent Concepts**: This is the set of concepts P that C is a subset of . For example, the concept 'Math' could be a parent of 'Linear Algebra'. If there is no suitable parent, a concept can be located beneath the 'mew'-concept - a specially designated concept with no skills associated to it and with no parents.
- **Connection-strength(s)**: For each parent $p \in P$, a concept C denotes a connection strength c_p from 0 to 1, specifying the degree of similarity or difference. All connection strengths collectively add up to 1, so that when an assessment needs to draw assessors from its parents, the connections strenghts can used to determine how many assessors should stem from each parent.
- **Expiration time**: Concepts can specify expiration times e_c to reflect that some skills become outdated, need to be maintained, or change over time. An example would be concepts related to taxation-laws, which are changed on a relatively frequent basis and where false or outdated information can lead to significant losses. Members who have been assessed in a concept longer ago than specified by the expiration time do not lose their certificates, but can no longer take part in the process of assessing others.
- **Members**: a set of individuals who have passed an assessment in the concept in question or in one of its children
- **Weights**: For each member the concept stores a set of weights, a positive integer and date, corresponding to the latest assessment in the concept. Weights are used to probabilistically call a member to act as assessors in assessments[1].
- **Owner**: An ethereum address that controls what data is saved on the concept and which can move the concept around by changing its parents.[3].

The Concept-Ontology

As all concepts have at least one parent, the entirety of all concepts forms a directed graph. The only concept without parents is the initial concept, the mew-concept. It does not present any particular subject or skill but serves as the root-node of the graph and as parent to all new concepts that are unrelated to the already existing ones. Thus, moving father away from the 'mew', concepts become more specific.

This network of relationship among knowledge-communities is valuable when sampling members to create a pool of potential assessors. If a concept has not enough members to create a pool of assessors of sufficient size, additional assessors will be drawn from the parent concepts.

As described in Governance & Upgrades the set of concepts and its definitions will be changing over time, with concept owners expected to update a concept's description and relation to parent concepts (in coordination with its members) and with members cloning and migrating their weights in case disagreements can not be resolved.

The Assessment Game

An assessment is the process by which a jury of qualified individuals (assessors) decides whether or not some candidate (assessee) fulfills the necessary conditions to become a member of a concept. When initiating an assessment in a concept, the assessee decides how many assessors they want and how much they are willing to pay to each one of them. That offer is forwarded to potential assessors (see *setup* for drawing specifics) who must stake the offered amount in order to accept. Thus, a market forms around assessments, allowing the system to scale from easy to assess, and hence cheap, concepts, to more involved, complicated, and hence expensive ones.

[1] Although it's possible to repeat an assessment, only the result of the most recent assessment will be taken into account for the weight.

[3] While this is not specified by the protocol, we intend the data-field (a bytes array) to be the ultimate source of truth of what a concept is about. As the owner is just any ethereum address, this allows for a variety of governance schemes to be implemented. The owner can also transfer ownership to another address or set it to zero, in which case the data-field becomes immutable.

Upon completion of the assessment, assessors are paid the price offered by the assessee and a proportion of their stake if they come to consensus around the applicants skill. The proportion of the stake being paid back is proportional to the assessor's proximity to the average score of the biggest cluster of scores and will also be added a portion of the stake of dissenting assessors, should there be any (see *payout* for details).

Also, the mechanism by which assessors log in their scores is designed such that colluding assessors can double cross each other, thereby creating a coordination problem in an adverserial environment, where the only point of coordination (schelling point) left is a truthful assessment. In case the majority vote of the assessors is positive, the assessed candidate will get i) a score in the assessed concept, similar to a grade in university or school, ii) a weight in the concept and iii) a weight in all parent-concepts, proportionally reduced by their respective connections strengths.

1.2.2 Assessment Process

A fathom assessment goes through five phases: A setup phase, where the assessors are called from the concept tree, the assessment, where the assessors determine the assessee's skill, commit- and reveal-phases, where the assessors log in their score and, at last, the calculation of the result.

For each phase this section is gonna depict the choices of the involved participants, their interactions and what happens if they deviate from the protocol.

Setup

Creating the assessment:

Wanting to be certified in a concept C, the assessee needs to specify the following parameters:

1. A time period during which they would like the assessment to start and end (latest start and end time).
2. The number of assessors N_a to be assessed by. While there is a minimum number of five assessors to guarantee a fair voting, the assessee might want to be assessed by a bigger number in order to receive a higher weight and higher chances to become assessors themselves (given that they end the assessment with a passing score).
3. The price $cost_a$ that each assessor will be paid.

Calling assessors from the concept-tree:

A pool of potential assessors is created by probabilistically drawing members from the concept and its parents. The selection of potential assessors happens according to a tournament-selection of size 2, starting at the assessed concept:

- Two members are picked at random and their weights are compared.
- The member with the higher weight is being added to the pool. In case of a draw, the member that was drawn first wins.

Thus each member has a chance of being called as assessor, whilst giving a higher chance to those with higher weights.

To make it hard to predict who will be in the pool, no more than half of the members of each concept can be called as assessors. Therefore, the maximum number of tournaments per concept c is limited by the number of members in the concept. Specifically Y is:

$$Y = \min(N_{req}, \frac{m_c}{2})$$

with N_{req} being the number of required assessors and m_c the number of members in the concept c.

After Y attempts, this selection process is repeated for each of the n_{p_c}'s parent concepts of c, using the connections strengths to the different parents to determine how many members are drawn from each parent.

The minimal size and the ideal size for the pool of assessor are subject to parameters and will grow with the amount of members in the network.

Assessors confirm by staking:

Each assessor that is being called, can decide to participate in it by staking the offered price. Once the desired number of assessors has confirmed, the assessment moves to the next stage. Assessors from the pool self-select whether they think would be competent judges on the concept in question. If so, they signal their intent to participate by staking the offered price. More considerations why assessors would or wouldn't want to confirm are elaborated in the incentive section. If not enough assessors can be found before the desired start-time of the assessment, the assessment is cancelled and everybody who deposited collateral is refunded.

Assessment of the candidate

In a fathom-assessment there is no notion or form what constitutes a test and the form or procedure of how candidates are evaluated is left to each individual assessor. Ultimately, assessors express their verdict of assessee's skill as a number on a scale (e.g. between 0 and 100) - with everything above half being considered a passing score.

Yet, what exactly defines a failing, passing or barely passing assessment can be different for each concept as well and should be agreed upon by the community. Moreover, the assessment could also be the place to put up some sybil protection mechanism in the form of extra requirements that make it hard to repeat an assessment (see sybil-attack for more details on how this could work).

Committing a Score

Sending in a score follows the commit-reveal procedure common in blockchain applications. Assessors signal that they have decided on a score by concatenating it with a secret element, also referred to as 'salt' and submitting its hashed value (hash=sha3(score+salt)).

If any assessors fails to commit a score before the assessment ends their stake is being burned. If, as a consequence, less assessors than would be required for the minimum size of a viable assessment have committed, the assessment is cancelled and everyone is being refunded. Otherwise, the assessment progresses to the next stage.

Steal and Reveal

To end the assessment, the assessors reveal their verdict by submitting their score and salt separately. Any assessor (or external person) who knows about another assessor's score and salt, can do so as well, thereby stealing half of the assessor's stake, burning the rest and eliminating him/her from the assessment game. This prevents the assessors from credibly guaranteeing each other their committment to logging in a specific score, thus making it harder to collude.

While stealing is possible at all times after an assessor has committed (even if others have not yet), revealing will only be possible after all assessors have committed and a buffer period of 12 hours has elapsed. The buffer ensures that there is time to challenge someone's commit, even if they waited until moments before the end of the assessment period to send it in. Should any assessor fail to reveal, their stake is burned and they are eliminated from the assessment game[2]. If the number of assessors decreases below the necessary minimum, the rest of the participants is being refunded and the assessment ends without a score.

[2] This should be unlikely, as at that point assessors have nothing to lose but rewards to gain.

Determining the Outcome

In order for an assessment to result in a final score, one score must be in consensus with enough other scores to form a 51% majority. Two scores are considered to be in consensus if their difference is less than the *consensus-distance* ϕ. If such a score s_{origin} exists, the final score s_{final} is computed as the average of all scores that are in consensus with s_{origin}.

Should there be two scores with majorities of equal sizes, the one that will result in a lower final score wins. If there is no point of consensus, the assessment is considered invalid and all stakes are burned. Otherwise the assessors' payments will be computed as described in the following section.

Also, in case the result s_{final} is a passing score, the assessee is registered as new member of the concept with a weight $w_i = s_{final} * N_{in}$.

Payout of Assessors

Payments to assessors consist of two parts, their returned stake and the assessee's reward. Both are attributed differently, depending on whether or not the assessor is inside out outside the majority cluster of winning assessors.

Therefore, an assessor i's distance $dist_i$ from the final score s_{final} is measured against the consensus range ϕ:

$$dist_i = \frac{|s_i - s_{final}|}{\phi}$$

Outside assessors ($1 < dist_i < 2$) only get back a part of their stake, reduced linearly in relation to their distance from the final cluster. Thus, an outside assessor j's payout is computed as:

$$payout_{out_j} = stake_j * \frac{(2 - dist_j)}{2}$$

Inside assessors ($dist_i <= 1$) get back their entire stake, any stake that is not returned to outside assessors (distributed equally among them) and a share of the assessee's reward, proportional to their proximity to the final score:

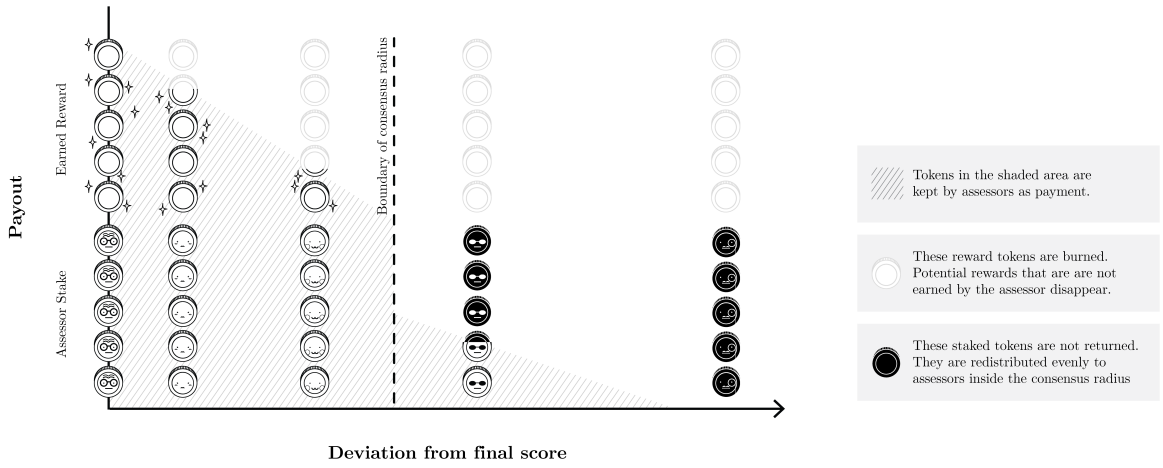
$$payout_{in_i} = stake_i + \frac{1}{N_{in}} * \sum_{j}^{N_{out}} (stake_j - payout_{out_j}) + r_a * (1 - dist_i)$$

,with N_{out} and N_{in} denoting the number of assessors outside and inside the winning cluster and r_a being the reward of the assessee.

Thus, the best case scenario for an assessor is to be inside the winning cluster, close to the final score, with a large minority outside of it.

Any payout that is not returned to the assessors will be burned (if it were redistributed, assessors could collude to cover a range of scores and redistribute amongst them without any loss).

This figure summarizes the payout mechanism in a single graph:



1.3 Concept Governance and Network Upgrades

This section introduces some of our considerations about how concepts will be defined and changed over time in a decentralized manner that leaves no single party with undue influence or decisionmaking power.

1.3.1 Concept Maintenance and Taxation

Getting right what it means to pass or fail a credential will be a crucial to allow assessors to come to consensus and is most likely very hard to get right on the first attempt.

While second-layer solutions like offchain-registries are a possible decentralized solution to this, we expect the data that is saved on the concept to be the source of truth about what the concept is about and what makes an assessee reach a pass or fail score.

Therefore, only the owner of the concept is allowed to change the concept's data field. While this can be used in a centralized manner it also allows for complicated governance schemes, e.g. if the owner were a smart contract that requires changes to the concept definition to be signed off by a majority of concept-members.

To incentivize the owners to keep a concept up-to-date and spend time and effort to source input from concept members and the wider community, concept owners can impose a 5% tax on all assessments run in the concept. That means that 5% of all network tokens that are spent on this concept will be redirected to an address of the owner's choice.

1.3.2 Resolving Conflicts by Cloning

In case there are conflicts about a concept-definition or the members feel like the concept-taxation is unjustified, there is the possibility to "clone" a concept.

Cloned concepts are just like regular concepts, i.e. all parameters can be freely defined, except that they additionaly designate an "original"-concept. Any members of the original concept can become members of the clone concept without an assessment simply by forsaking their membership in the original.

'Migrating' members will have the same weight in the clone that they had in the original.

Thus owner are incentivized to only tax members if they actually provide value to the community and unresolvable disagreements about credential-content can be resolved by opting out of the definition without further costs.

1.3.3 Network Upgrades

The cloning mechanism can also be used to issue protocol updates such as changing the assessment-parameters or the assessor calling mechanism. To roll out such a change, a new network with a special set of cloning-rules would be deployed. The new network will implement the updated protocol **and** allow concepts to be clones of concept in the old-network.

Therefore, at any point in time, anyone can just deploy a new version of the fathom-network and convince the community to adopt it by migrating their memberships to the relevant concepts in the new network. To increase transparency and allow community involvement, the first network will implement a time-delay before changes can come into effect.

1.4 Security

The fathom protocol derives security from both it's technical implementation and incentive structure. There are specific attack vectors that it mitigates against as well as some areas for further research.

1.4.1 Incentive Structure

While traditional credentials are meaningful because they are backed by reputable institutions, a fathom credential is meaningful because it is the result of many individuals having undergone a financial risk in the assessment game in order to create it. This section will lay out the decisions fathom users face when participating in that game, as well as the economic risks associated to them. Specifically, we show i) when assessors are likely to participate in an assessment in the first place and ii) why they can not collude with each other in order to shortcut the work associated with a truthful assessment.

Incentives for members to confirm or decline an assessment

This case is especially relevant for creating new concepts - as those will be initially empty and rely on members of its parent to participate in as assessors. A member of a concept who has received an offer to be an assessors will consider whether...

- They feel competent enough in their abilities to come to the same conclusion as a group of other, randomly selected assessors that are confident in *their* abilities.
- The concept in question is well enough defined so that assessors with similar impressions of the assessee's skill will be able to translate these into similar scores.

Incentives for assessors to grade truthfully

As truthfully assessing someone requires effort and the assessors payout is pegged to their alignment with each other, there is a motivation for them to collude, e.g. by agreeing ahead of time which score to commit. The creation of an adverserial environment between assessors is thus vital for the protocol to function as intended. Therefore, several mechanisms are put in place: First, assessors are paid out more if there are dissenting assessors (see Figure 1) Consequently, any assessor taking part in a collusion of X assessors, must be afraid that they will be double crossed by a subgroup of more than $X/2$ assessors. These are motivated to do so because they would be rewarded with part of the crossed assessors stake. Moreover, it is not possible for assessors to credibly prove to another assessor that they have actually committed to a collusion and logged in a previously agreed-upon score. In order to do so, the proving party would have to reveal their score and salt to the other assessors. Yet with this information, the other assessors could simply steal the assessors stake, which would eliminate the former assessor from the assessment and directly transfer the half of the revealed assessor's stake to the revealer.

1.4.2 Attack Vectors

This section will outline some of the general classes of attacks against the protocol and a subjective view of their complexity, severity and to what degree they are considered to be mitigated.

Sybil Attacks

In a sybil attack, the attacker creates many false identities and then uses them to subvert the system, e.g. by controlling most of the identities in a concept, giving him control over who will be accepted and the ability to create assessments for himself in order to steal the stakes of other assessors.

To set up such an attack the attacker would, instead of being assessed by many assessors in one assessment, create multiple assessments with fewer assessors. This would be the same amount of work but result in four identities in the concept. Repeating the procedure, the attacker could count on some of his identities being called as assessors in which case the subsequent repetitions would become cheaper and less time-consuming until they have the majority in the concept or are called multiple times as assessor such that they can set up a 51% attack on individual assessments. In such a scenario, the attacker could control the outcome of the assessment and steal the stake of the other assessors.

Severity of attack: While a sybil attack does cost a fair amount of money to set up, the potential benefits are big enough to incentivize a try. As a compromised concept can potentially 'poison' its parent concepts as well and thus potentially effect the entire tree, we consider it to very severe.

Complexity: While a sybil attack is fairly complex, it can be effectuated by a single attacker, which is why it would be careless to assume that the degree of complexity will be a deterrent factor.

Degree of Protection: One possible mitigation that is not yet part of the protocol, will be to split the certificate and the right to be an assessor in two separate assessments. While this does not address the fundamental issue, it makes it easier for the sybil-protection measures to be integrated into the assessment process. For example, the assessment to become an assessor could ask the to-be-assessors for some piece of their own work or something that is new and can not be readily found on the internet as would be the case with the mere knowledge or skill required in the concept.

Simple Trolling

A troll, for arbitrary reasons, might try to poison the fathom network by creating a bunch of bogus assessments or concepts or by behaving irrationaly while being an assessor. In all cases, such behavior is expensive and ineffective, as his stakes are burned (when not following through with an assessment) or redistributed to others (when logging in bogus scores).

Bogus concepts will simply incur costs on the troll and be filtered out by assessors (see *incentives*). Creating bogus assessments as assessee will be even more costly (transaction costs and the fees for assessors). The worst effect a troll can have is to become an assessors and to prematurely end the assessment, if as a consequence of their behavior, its size is reduced below the minimum of five. In that case all other participants will be refunded, though.

Complexity: Behaving irrationaly is simple and so is attacking the system this way.

Severity: With no financial costs to other participants these kinds of attacks are not considered severe. An exception might be the creation of concepts, which if done by a well-resourced attacker, amounts to spamming the system.

Degree of protection: We consider simple trolling to be sufficiently discouraged because of the associated costs. If such behavior would be escalated into a spam-attack of greater proportions, the degree of protection will depend on the users or the fathom frontends ability to filter concepts and assessments by meaningful criteria.

P + epsilon attack

In a P + epsilon attack, the attacker circumvents the incentivization by creating a mechanism that others can trust in because it gives them a credible guarantee about the attacker's behavior. While this would have been difficult in a

pre-blockchain era, smart contracts are nearly ideally suited to implement such mechanisms.

The attack works like this: In a schelling-point game, the assessors are being paid out the same amount P. regardless of the result (option A, B, C or any other...). The attacker, let's say wanting to push for a certain option A, will credibly guarantee anyone voting for A that he will be paid P+ϵ, if they vote A and the majority doesn't. Assuming a system that is not dominated by altruistic actors, voting A is now the game-theoretically best option (guaranteed maximal payout). Therefore, the majority will vote A and the attacker will have taken over the mechanism - at zero cost.

Although there exist some protection mechanisms that can increase the attackers risk (size of the needed bribe) and some counter-coordination mechanisms that come close to defeating such an attack, there is currently no guaranteed countermeasure.

Complexity: As the crucial element of this attack is the mechanism by which the attacker commits to his intention to paying out in case the bribed voter is not in the majority, the complexity is proportional to the difficulty of construing such a mechanism. In the case of fathom, the difficulty to reconstruct the relevant information (did an assessor really vote for the desired option A?). Currently, this is rather simple, so setting up this attack would not be very complex.

Severity: As this attack can disrupt the system at potentially zero-cost, we consider it to be very severe.

Degree of Protection: As of right now, the protocol is not protected against such measurs. Future versions of it could implement some more complicated schemes in order to keep the scores of individual assessors secret and make it harder to retrieve the individual assessors' scores.

1.5 Vision

We believe that a participatory protocol tied directly to the communities practicing skills and defining ideas will diminish the gap between credentials that can be credibly assessed and issued and the wide variety of skills and abilities that people are capable of.

Tying economic incentives to this social process and ontology, such that they are both visible to everybody and aligned amongst all those participating, allows for fathom-credentials to be trustworthy and transparent.

Distributing the work required to communities allows the system to scale and be accessible to anyone, no matter their previous records, achievements or socio-economic circumstances.

We believe that through these traits fathom enables a world where people are free to shape their own experiences, communicate them to others, and organize to achive shared ambitions.

1.6 Glossary

Assessor A user who validates whether one can be a member of a concept or not.

Assessee The subject of an assessment who wants to earn a credential and join a `concept`.

Staking The process by which assessors confirm for assessments by locking up a fixed amount of tokens. These tokens are either returned to them or burned, depending on their performance assessing.

Concept Represents the shared knowledge, skillset, or other attribute that the users who have attained the concept have in commom. Examples could range from Calculus proficiency to English fluency to marathon completion.

Weight Upon a successful assessment an individual earns a weight in a concept. This is calculated as a function of their score, the number of assessors in the largest cluster, and the time the assessment was taken.

Index