The Body of Knowledge – an Introduction

John McDermid

29 June 2018

UNIVERSITY of York

ASSURING AUTONOMY
INTERNATIONAL PROGRAMME

1. Introduction and Context

The Assuring Autonomy International Programme was established in January 2018, with the aim of enabling the benefits of robotics and autonomous systems (RAS) to be realised, by facilitating their safe, assured, and regulated adoption. To do this, the Programme needs to provide material to support the assurance and regulatory processes, but also inform the development of RAS, since RAS assurance is expected to require the use of appropriate design principles in their development, e.g. to enable reasoning about the capabilities of systems employing machine learning. The initial Programme documentation referred to a framework for assurance and regulation; the Programme now uses the term Body of Knowledge (BoK) to reflect the need to address design issues, alongside regulation and assurance.

The BoK will be one of the main tools enabling the Programme to meet its aim. As such, the BoK will need to be very wide reaching. The BoK will be a structured, curated knowledge base covering state-of-the-art techniques, methods and processes for engineering, assuring, deploying, regulating and operating RAS in multiple domains including maritime, air and healthcare. Developing such a BoK is challenging, and it is anticipated that this will be an ongoing activity for the life of the Programme – 5 years – or more. The aim of this document is to set out the objectives for the BoK, explain how it is expected to develop and evolve, and to set the context for more detailed content.

If successful, the BoK will be the *de facto* reference source for those wanting to assure and regulate RAS – and thus for developers who wish to reduce the risks that their systems are unsafe, or cannot be approved for operation. The Programme is intended to address the full range of application of RAS, and to be truly international, addressing the use of RAS across the globe. This potentially makes the BoK very large, and very complex. Thus the BoK will only be successful if it focuses on critical issues that cut across domains of application and technologies, and provides information that can be adopted and adapted by developers and regulators, for their own domains. By focusing the BoK will say less, but be more useful!

At this stage, the fledgling BoK is presented as a set of documents. In time, it will be a globally accessible knowledge base (perhaps best thought of as a trustworthy wiki) that can be queried by different stakeholders, to find information pertinent to them, e.g. what set of techniques can be used for assuring human-robot interaction, and what are the associated regulations in assistive (social care) robotics, if any. It is likely that a version of this document will be made available to serve as an introduction to the BoK for new users (like a “read me” but giving more of the motivation for the BoK structure and content).

This document covers the:

- Scope and content of the BoK – the intended content of the BoK; it will be technically focused but also address some of the wider socio-technical issues, e.g. risk acceptance;
- Stakeholders for the BoK – the organisations and classes of individual expected to use the BoK (as consumers and contributors) and the criteria that they might use for evaluating the utility of the BoK;
- Development and evolution of the BoK – the way in which the Programme plans to develop and evolve the BoK, recognising that these are ongoing tasks.

It is expected that the foundations for the BoK will be refined over time, and part of the reason for making this document publicly available is to get feedback to assist in the refinement process.

At the time of writing (June 2018) two companion documents are being produced: an initial discussion of problems and principles and a State of the Art (SotA) in the automotive domain. These two documents are on a more limited circulation, in order to get expert feedback and input before disseminating them more widely.

2. Scope and Content of the BoK

The framework will be technically focused, and will contain a set of definitions, principles, practices and criteria[1] covering:

- Product technology – technology specific issues, which will include enabling technology, e.g. processors and communications networks, and this should address failure modes for machine learning technologies, etc.
- Product – domain/function-specific information focusing on risk management (design), assurance and regulation around the product breakdown into sensing, understanding, decision-making and actuation (SUDA);
- Process – covering design, assurance/regulation and operation and focusing on safety (and the impact of cyber security on safety), recognising that safety needs to be considered as much more of a through-life issue.

A distinction is made between product technology, e.g. a Lidar, and a product that might use a Lidar in a particular domain. This distinction is believed to be useful as there will be general properties of technologies that it is useful to record, but full understanding of the capabilities and limitations of the technology will depend on its context of use.

The BoK will use a canonical model of the capabilities of a RAS, representing a product breakdown into sensing, understanding, decision-making and actuation, which we refer to as SUDA. The core technical part of the BoK will focus on the SUDA elements, but will also address support technology such as communications and processors.

The core material will be supported by socio-technical material on public understanding and acceptance of risk, and on ethical issues, as appropriate.

As the potential scope of the BoK is very wide it is necessary to focus on issues that can significantly impact the ability to assure and regulate RAS, which we refer to as critical barriers to assurance and regulation (C-BARS). The consequences of failing to "overcome" these barriers, for individual RAS, might be:

- A safe system cannot be deployed (losing benefit);
- An unsafe system is deployed (as it is approved due to lack of contrary evidence).

For a whole domain, e.g. autonomous driving or healthcare, the consequences of unsolved barriers are potentially:

- A low rate of adoption of safe technology;
- A high level of accidents and incidents leading to a backlash.

The Programme will seek to identify C-BARs on an ongoing basis (see the discussion of the BoK evolution below). In the immediate term many C-BARs will be easily identified since:

- They inhibit the assurance of RAS, and
- There is a critical gap in the knowledge in the area (both in academia and industry)

These can be thought of as "challenge problems" (CPs); over time CPs will be solved but they will remain barriers, and approaches to overcoming the C-BARs will be key elements of the BoK.

There content will represent three related aspects of the CPs or C-BARs:

- Problems – the C-BARs or CPs to be addressed in designing, assuring and regulating RAS, e.g. validation of machine learning algorithms, understanding of sensor limitations, and assessment of risk in the presence of uncertainty;

[1] It is envisaged that the BoK will cover methods, and that tools will be referenced from the BoK.

- Principles – general strategies for designing, assuring and regulating RAS, e.g. adapting machine learning so it exports a representation of the learnt rules, use of command-monitor architectures;
- Practices – particular techniques for designing, assuring and regulating RAS, e.g. architectures, test methods, approaches to assurance arguments.

It should be noted that the BoK is not a safety case or assurance case; however it is expected that it will be used to guide the production of assurance cases, e.g. by identifying the SotA for assuring some aspect of a design, e.g. object classification in machine vision.

3. Development and Evolution of the BoK

The BoK cannot be static. For example, if a new solution to an important problem, e.g. the validation of sense and avoid algorithms for autonomous air vehicles, is produced, then the BoK needs to be updated to reflect this. However there is a need for quality control over the BoK – it cannot just accrete information – there must be a curation process that ensures that what is included is valid and up to date. Thus, in the steady state, the BoK will be fairly stable in scope and structure, but the content will need to evolve to ensure that it is up to date; it will be important to ensure governance of the evolution, but the initial development is more challenging!

In order to develop the BoK the Programme must identify CPs/ C-BARs, and then stimulate work to solve the problems/overcome the barriers. This is best done collaboratively, as the Programme does not have the "monopoly of wisdom" on these issues, and there is already considerable relevant work going on in industry and academia. The intent is to use three related mechanisms to define (and later refine) the CPs/ C-BARs:

- Stakeholder[2] workshops – to consider both scoping of the BoK, and identifying specific CPs/ C-BARs that are focused enough it is possible to tell if they have been solved (or ameliorated) but general enough that they are applicable across a range of domains and systems;
- Liaison with other programmes/activities – work with other initiatives, e.g. the 3Ai[3] programme in Australia, and the Safety of Autonomous Systems Working Group[4] (SASWG) in the UK, who are addressing similar issues;
- Elicitation of knowledge from experts – drawing on the Programme's international community, and Visiting Fellows[5], to provide deep knowledge on particular problems in the experts' domains.

These are related as, for example, someone from a linked programme might become a Visiting Fellow, and could contribute to organising and running a workshop.

Scoping workshops may be run internally to the Programme, but work on identifying CPs/ C-BARs will be conducted in collaboration with the international community. An initial set of CPs/ C-BARs has already been identified[6], and it is intended that a series of workshops will be held during the second half of 2018 to produce a more complete and balanced set of problems to guide the remaining four years of the Programme. Again it is not expected that this set will be static, but it is expected that it will evolve more slowly over time.

There is another important mechanism for defining and solving CPs/ C-BARs and contributing to the BoK. The Programme is supporting a set of Demonstrator projects, prototyping or deploying RAS, by

[2] An initial list of stakeholders is included at Annex 1.

[3] See: https://cecs.anu.edu.au/3a-institute

[4] See: https://scsc.uk/ga

[5] A programme of visiting fellowships is being instigated in the second half of 2018.

[6] See Annex 2 for an example C-BAR/CP.

funding associated work on assurance and/or regulation of the RAS. The Demonstrators will be one of the key users of the BoK (see Annex 1) but will also contribute to the BoK. Further, in terms of the formation of the BoK, proposals for Demonstrators will be invited to propose new CPs/CARBs that need to be addressed by the BoK, as well as responding to those that have already been identified.

It is important that the BoK is widely accepted, and is subject to evaluation. This will be facilitated through making it accessible on-line, but will also be supported through more “formal” events. The Programme will run annual conferences involving Demonstrators and other stakeholders; these will be used as a forum for promulgating changes to the BoK, and for soliciting feedback on the content.

In time, the BoK must be accessible internationally, through the Internet, and able to be queried in order to find appropriate information. For example, it is intended to support queries such as “what are the problems of validation of Lidars?” and “what are the SotA practices for validation of Lidars used for sensing in the maritime environment?”.

The BoK also needs to be actively managed (curated) to ensure its validity and currency. This is likely to be achieved via a form of “managed wiki”, perhaps similar to the system used by the National Institute for Clinical Excellence (NICE). As work is done on the scope, structure, and initial content of the BoK during the second half of 2018, work will also be undertaken to identify appropriate technical infrastructure and governance mechanisms for developing, evolving and disseminating the BoK.

4. Concluding Remarks

The BoK will be a key, evolving, deliverable for the Programme, and the quality of the BoK will be a critical success factor for the Programme. This document has set out the current thoughts in the Programme on the scope of the BoK, and the viewpoints and stakeholders that the BoK needs to support in order to be effective. The reason for issuing this document, prior to developing the “BoK proper” is to get community feedback on the conceptualisation for the BoK both to improve it and to get “buy-in” to collaborative development of the BoK.

Annex 1: Stakeholders

The following is the current view of Stakeholders for the BoK, and is subject to revision.

Table 1: Stakeholders and Evaluation Criteria

Stakeholder	**BoK Evaluation Criteria (the BoK provides …)**
System Developers	Design constraints and assurance problems, principles and practices (including the SotA) relevant to the domain, function and technology of their system Guidance on safe and secure use of product technologies relevant to the domain, system, function and technology of their system Regulatory principles, practices and standards relevant to their domain
Regulators/ Assessors	Guidance on how to assure or regulate product technologies or systems in their domain, and information on best practice from other domains to use as comparators Guidance on risk acceptance
Policy Makers	Definition of problems and principles relevant to their domain which, if not addressed by policy, could impede the safe take-up of RAS in their domain of concern Information on public perception of risk and benefits of RAS relevant to their domain
Demonstrator Leads	SotA relating to design, assurance and regulation of RAS in their domain, on which the Demonstrator can build Identifying gaps in knowledge which the Demonstrator can seek to fill (and contribute to the BoK)
Research Community	SotA relating to design, assurance and regulation of RAS and underlying technologies Identifying gaps in knowledge that can usefully be addressed
Operators	Operational risks (residual uncertainties) arising from the use of their system, given its domain, function and technology Approaches to monitoring and conducting incident/accident analysis relevant to their system, given its domain, function and technology
Standards Bodies	Information on the SotA relevant to product technology and domains to inform the standards development process
Trainers	Design constraints and assurance problems, principles and practices (including the SotA) relevant to a domain or domains, system function and technology Guidance on safe and secure use of product technologies relevant to a domain or domains, system, function and technology Regulatory principles, practices and standards relevant to a domain or domains
Insurers	Understanding of responsibility for accidents Assessment of risk of systems in operation

Lawyers	Understanding of responsibility for accidents Understanding of ethical constraints on design and operation
Public	Confidence that the professional community has access to the SotA relevant to their technology and domains Clear understanding of risks and benefits of RAS about which they are concerned, e.g. because they might be affected by them

Annex 2: Example Critical Barrier to Assurance and Regulation

The following example is for illustrative purposes only. As well as illustrating a C-BAR, it shows how the problem/principle/practice distinctions in the BoK might be interpreted in defining a demonstrator project.

Assurance of Human-Robot-Interaction in Social Care

This demonstrator uses humanoid robots to care for the elderly and infirm in a domestic environment. These robots will interact physically with people, enabling them to stand and walk, or to carry them, see Figure 1. The robots are capable of lifting weights up to 100kg, thus they can pose a risk of harm to the individual.

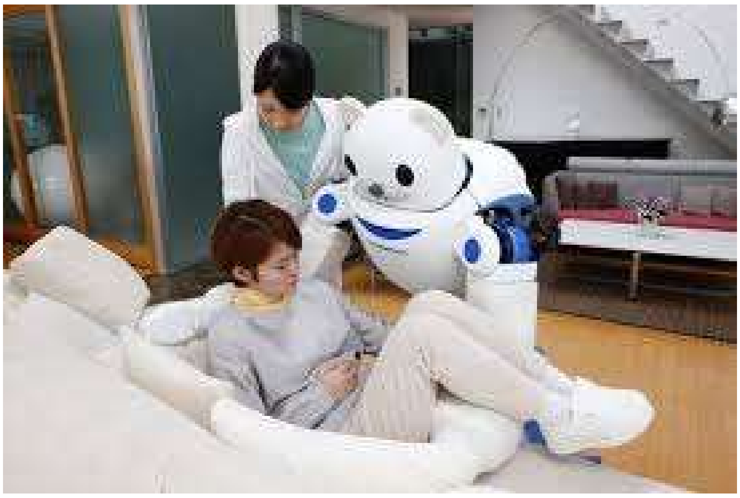

Figure 1: Robot lifting a teenager

The challenge is:

- To demonstrate that a robot cannot interact with the human they are supporting in a way that will cause harm, including bruising.

The principles to be adopted are:

- The robot assesses manoeuvres, and doesn't attempt them if unsafe, e.g. a lift would interact with a bandaged limb;
- The force applied is distributed, so pressure never exceeds a given threshold;
- Impact velocity is always below a specified threshold.

The practices to be explored and validated are:

- Use of simulation to compute maximum forces in challenging manoeuvres, for representative physiological types;
- Validation of simulation through use of dummies with pressure and impact sensors;
- Trials with humans, including "emergency" response, e.g. trips.

The demonstrator would deliver a method for assessing safe physical interaction with humans, and a data set from the trials that might assist other projects.